

OM of: US-09-303-518d-651 to: SwissProt\_40.\* out\_format: pfs  
 Date: Jun 30, 2002 8:57 AM

About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODE=frame-n2p,model -DEV-xlh  
 -Q/cgn2.1/USPTO\_spool/US09303518/runat\_28062002.142714.4362/app\_query.fasta.1.23501  
 -DB=SwissProt\_40 -QFM=fastan -SUFFIX=rs -GAP=12.000  
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000  
 -OGAP=4.500 -OGAPEXT=0.050 -XGAP=10.000 -XGAPEXT=0.500  
 -RAPP=6.000 -RGAPEXT=7.000 -YGAPO=10.000 -YGAPEXT=0.500  
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62  
 -TRANS=human40.cdi -LIST=100 -DOCALLIGN=200 -THR\_SCORE=pct  
 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs  
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
 -USR=US09303518.@CGN1.1.440 -NCPU=6 -ICPU=3 -LONGLOG  
 -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPPX -WAIT -THREADS=1

## Search information block:

Query: US-09-303-518d-651  
 Query length: 4350  
 Database: SwissProt\_40.\*  
 Database sequences: 105324  
 Database length: 38719550  
 Search time (sec): 217.960000

## score list:

Sequence	Strid	Orig	ZScore	EScore	len	Documentation
SwissProt_40:1GA3_HAEIN +	1554.50	1563.55	1.4e-80	1545	P45385 haemophilus influenzae	
SwissProt_40:1GA1_HAEIN +	1536.50	1545.34	1.4e-79	1541	P42782 haemophilus influenzae	
SwissProt_40:1GA0_HAEIN +	1519.00	1526.81	1.4e-78	1594	P44699 haemophilus influenzae	
SwissProt_40:1GA2_HAEIN +	1517.00	1524.74	1.8e-78	1702	P45384 haemophilus influenzae	
SwissProt_40:1GA1_NEIGO +	1508.00	1516.52	5.8e-78	1532	P09790 neisseria gonorrhoeae	
SwissProt_40:1GA4_HAEIN +	1475.00	1481.50	4.3e-76	1849	P45386 haemophilus influenzae	
SwissProt_40:YB7A_ECOLI +	372.50	366.23	6.7e-14	1569	P57143 escherichia coli. hyphimurium	
SwissProt_40:YB7A_ECOLI +	264.50	259.98	9.5e-08	1250	P25927 salmonella typhimurium	
SwissProt_40:YB7A_ECOLI +	263.50	257.76	9.3e-08	1250	P45508 escherichia coli. hyphimurium	
SwissProt_40:YB7A_ECOLI +	244.50	241.66	1.1e-06	863	P33924 escherichia coli. hyphimurium	
SwissProt_40:YB7A_ECOLI +	242.00	231.98	1.6e-06	2003	P33666 escherichia coli. hyphimurium	
SwissProt_40:YB7A_ECOLI +	241.50	230.48	1.7e-06	2249	P15921 rickettsia rickettsiae	
SwissProt_40:YB7A_ECOLI +	236.00	229.41	3.3e-06	1325	P33051 escherichia coli. hyphimurium	
SwissProt_40:YB7A_ECOLI +	234.00	223.80	4.5e-06	2021	P05657 rickettsia conorii	
SwissProt_40:YB7A_ECOLI +	229.50	226.02	7.5e-06	910	P14283 bordetella pertussis	
SwissProt_40:YB7A_ECOLI +	227.00	223.38	1.0e-05	922	P24328 bordetella pertussis	
SwissProt_40:YB7A_ECOLI +	224.50	219.83	1.4e-05	1039	P33180 escherichia coli. atypica	
SwissProt_40:YB7A_ECOLI +	224.00	220.44	1.5e-05	911	P003035 bordetella bronchiseptica	
SwissProt_40:YB7A_ECOLI +	217.00	208.69	4.0e-05	1577	P14666 proteus mirabilis	
SwissProt_40:YB7A_ECOLI +	214.50	207.89	5.4e-05	1286	P03155 escherichia coli. adhesion	
SwissProt_40:YB7A_ECOLI +	206.00	198.76	0.0002	1367	P08640 saccharomyces cerevisiae	
SwissProt_40:YB7A_ECOLI +	205.00	196.17	0.0002	1645	P06989 r outer membrane protein	
SwissProt_40:YB7A_ECOLI +	203.00	194.10	0.0002	1655	P09K43 r outer membrane protein	
SwissProt_40:YB7A_ECOLI +	201.00	196.99	0.0003	928	P45354 haemophilus influenzae	
SwissProt_40:YB7A_ECOLI +	200.00	190.21	0.0004	1829	P55127 neisseria meningitidis	
SwissProt_40:YB7A_ECOLI +	198.50	189.54	0.0004	1654	P03047 r outer membrane protein	
SwissProt_40:YB7A_ECOLI +	197.00	186.04	0.0005	2090	P33658 homo sapiens (human)	
SwissProt_40:YB7A_ECOLI +	193.00	183.12	0.0009	1829	P09195 neisseria meningitidis	
SwissProt_40:YB7A_ECOLI +	191.50	186.10	0.0011	1077	P002308 drosophila melanogaster	
SwissProt_40:YB7A_ECOLI +	191.50	182.11	0.0011	1723	P09812 chlamydia pneumoniae	
SwissProt_40:YB7A_ECOLI +	191.00	181.94	0.0012	1656	P06653 r outer membrane protein	
SwissProt_40:YB7A_ECOLI +	190.50	181.50	0.0012	1643	P05320 r outer membrane protein	
SwissProt_40:YB7A_ECOLI +	186.00	180.05	0.0022	1140	P04893 saccharomyces cerevisiae	
SwissProt_40:YB7A_ECOLI +	185.50	176.61	0.0024	1609	P09605 chlamydia pneumoniae	
SwissProt_40:YB7A_ECOLI +	185.00	177.92	0.0025	1300	P14914 rickettsia rickettsiae	
SwissProt_40:YB7A_ECOLI +	184.00	171.93	0.0030	2334	P00783 bacillus subtilis	
SwissProt_40:YB7A_ECOLI +	183.50	176.13	0.0030	1341	P07067 bacteriophage t2	
SwissProt_40:YB7A_ECOLI +	183.00	180.05	0.0031	797	P28968 equine herpesvirus	
SwissProt_40:YB7A_ECOLI +	183.00	174.08	0.0033	1609	P23553 saccharomyces cerevisiae	
SwissProt_40:YB7A_ECOLI +	182.50	173.58	0.0035	1608	P15320 serratia marcescens	

SwissProt_40:P3P_LACIC +	181.50	171.14	0.0041	1902	P15292 lactococcus lactis
SwissProt_40:P3P_LACIC +	180.50	164.72	0.0049	3591	P12255 bordetella pertus
SwissProt_40:P3P_LACIC +	178.50	169.19	0.0059	1672	P09192 chlamydia muridar
SwissProt_40:P3P_LACIC +	178.00	170.89	0.0062	1290	P55981 heliobacter pylori
SwissProt_40:P3P_LACIC +	177.50	172.18	0.0065	1045	P09489 serratia marcesc
SwissProt_40:P3P_LACIC +	177.00	171.01	0.0070	1130	P09747 mus musculus (mou)
SwissProt_40:P3P_LACIC +	176.50	173.15	0.0072	827	P25062 halobacterium volc
SwissProt_40:P3P_LACIC +	176.50	170.54	0.0074	1125	P27546 mus musculus (mou)
SwissProt_40:P3P_LACIC +	176.00	168.86	0.0080	1291	P04825 heliobacter pylori
SwissProt_40:P3P_LACIC +	175.50	165.06	0.0089	1902	P16271 lactococcus lactis
SwissProt_40:P3P_LACIC +	175.00	167.81	0.0091	1296	P04847 heliobacter pylori
SwissProt_40:P3P_LACIC +	173.00	167.03	0.0117	1120	P07607 escherichia coli
SwissProt_40:P3P_LACIC +	173.00	162.53	0.0123	1902	P02470 lactobacillus parv
SwissProt_40:P3P_LACIC +	172.50	165.41	0.0126	1276	P09289 chlamydia pneumon
SwissProt_40:P3P_LACIC +	171.50	166.83	0.0140	959	P002629 saccharomyces cer
SwissProt_40:P3P_LACIC +	171.50	165.29	0.0143	1150	P12021 sus scrofa (pig) . f
SwissProt_40:P3P_LACIC +	169.50	171.48	0.0169	437	P21561 haloterrax sp. (st)
SwissProt_40:P3P_LACIC +	168.00	163.49	0.0221	936	P07591 citrobacter freund
SwissProt_40:P3P_LACIC +	168.00	160.46	0.0228	1337	P33653 streptococcus do
SwissProt_40:P3P_LACIC +	167.00	161.78	0.0253	1016	P08480 chlamydia trachom
SwissProt_40:P3P_LACIC +	167.00	158.70	0.0261	1460	P09311 chlamydia muridar
SwissProt_40:P3P_LACIC +	166.50	163.41	0.0264	790	P24856 nototheria corlicer
SwissProt_40:P3P_LACIC +	166.00	155.44	0.0305	1902	P15293 lactococcus lactis
SwissProt_40:P3P_LACIC +	165.50	158.25	0.0314	1287	P04825 heliobacter pylori
SwissProt_40:P3P_LACIC +	165.00	160.74	0.0325	905	P44602 haemophilus influe
SwissProt_40:P3P_LACIC +	164.50	158.47	0.0353	1113	P002630 saccharomyces cer
SwissProt_40:P3P_LACIC +	164.00	158.66	0.0374	1026	P03744 bacteriophage t4 .
SwissProt_40:P3P_LACIC +	164.00	156.61	0.0382	1306	P32334 saccharomyces cer
SwissProt_40:P3P_LACIC +	163.50	152.11	0.0425	2090	P5161 mesocricetus aur
SwissProt_40:P3P_LACIC +	163.00	160.04	0.0415	774	P03764 bacteriophage lamb
SwissProt_40:P3P_LACIC +	163.00	154.64	0.0440	1462	P49331 streptococcus mu
SwissProt_40:P3P_LACIC +	162.00	152.08	0.0509	1754	P08418 chlamydia trachom
SwissProt_40:P3P_LACIC +	162.00	152.00	0.0509	1770	P08419 chlamydia trachom
SwissProt_40:P3P_LACIC +	161.50	154.80	0.0525	1199	P25591 rattus norvegicus
SwissProt_40:P3P_LACIC +	161.00	156.57	0.0547	917	P45355 hemophilus influe
SwissProt_40:P3P_LACIC +	161.00	150.83	0.0581	1802	P41809 saccharomyces cer
SwissProt_40:P3P_LACIC +	160.50	153.28	0.0601	1273	P38968 saccharomyces cer
SwissProt_40:P3P_LACIC +	160.00	154.72	0.0628	1013	P08489 chlamydia trachom
SwissProt_40:P3P_LACIC +	160.00	154.23	0.0631	1072	P36225 bos taurus (bovin
SwissProt_40:P3P_LACIC +	160.00	150.75	0.0655	1616	P32805 bacillus circulac
SwissProt_40:P3P_LACIC +	159.50	153.94	0.0672	1045	P28805 serratia marcesc
SwissProt_40:P3P_LACIC +	159.00	150.84	0.0737	1419	P10368 candida albicans
SwissProt_40:P3P_LACIC +	158.50	156.76	0.0735	666	P39046 enterococcus hirac
SwissProt_40:P3P_LACIC +	158.50	154.03	0.0756	918	P34487 caenorhabditis el
SwissProt_40:P3P_LACIC +	158.00	149.46	0.0843	1597	P11001 streptococcus do
SwissProt_40:P3P_LACIC +	158.00	149.33	0.0795	1481	P38939 t amylopullulanase
SwissProt_40:P3P_LACIC +	158.00	148.85	0.0848	1592	P027470 streptococcus pyo
SwissProt_40:P3P_LACIC +	157.00	149.64	0.0947	1288	P09K45 heliobacter pylori
SwissProt_40:P3P_LACIC +	157.00	142.90	0.1018	2845	P06135 mus musculus (mou)
SwissProt_40:P3P_LACIC +	156.50	149.04	0.1012	1302	P09K09 neisseria mening
SwissProt_40:P3P_LACIC +	156.00	149.51	0.1069	1161	P28284 herpes simplex vir
SwissProt_40:P3P_LACIC +	155.50	148.52	0.1146	1229	P09223 homo sapiens (hum
SwissProt_40:P3P_LACIC +	154.50	143.99	0.1355	1861	P38536 t amylopullulanase
SwissProt_40:P3P_LACIC +	154.00	148.96	0.1364	973	P08419 chlamydia trachom
SwissProt_40:P3P_LACIC +	154.00	148.60	0.1370	1018	P34333 caenorhabditis el
SwissProt_40:P3P_LACIC +	154.00	145.69	0.1413	1433	P16397 bacillus subtilis
SwissProt_40:P3P_LACIC +	153.00	147.58	0.1560	1018	P14738 staphylococcus au

seq\_name: SwissProt\_40:1GA3\_HAEIN

seq\_documentation\_block: PRT; 1545 AA.

AC P4585: STANDARD;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL1 protease).

GN IGA.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

NCBI\_TaxID=727;

```

RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-HK393 / NCTC 8467 / SEROTYPE B;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.:
RT "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases.";
RL J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR: CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING IMPACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
CC certain Pro-1-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M87490; AAA24967.1;
DR MEMOPS: S06.001;
DR InterPro: IPR000710; IGA_S6.
DR Pfam: PF02395; IGA1;
DR PRINTS: PR00921; IGASERPTASE.
KW Hydrolyase; Serine protease; Transmembrane; zymogen; signal.
FT SIGNAL 1 25
FT CHAIN 26 1012
FT PROPEP 1013 1545
FT ACT_SITE 292 292
FT ACT_SITE 1545 AA; 170627 MW; 3EDD753988F6D478 CRC64;
SQ SEQUENCE 1545 AA; 170627 MW; 3EDD753988F6D478 CRC64;

alignment_scores:
Quality: 1554.50 Length: 1724
Ratio: 1.699 Gaps: 53
Percent Similarity: 53.074 Percent Identity: 26.798

alignment_block:
US-09-303-518D-651 x IGA3_HAEIN ..
Align seg 1/1 to: IGA3_HAEIN from: 1 to: 1545

64 CGCTTCGCGCTGCTTACTGACCATATGCTGCTGCTGCGCATCTTCC 113
   :::::::::::::::::::::
5 LysPheLysLeuAsnPhelIeAlaLeuThrValAlaIleuThr 21
   :::::::::::::::::::::
114 CCAGCTTGGCGCGGACACTTATTCGCGCATCACTACCAATPACTATC 163
   :::::::::::::::::::::
21 cTtThGlAlaIaLeuValaIgaSPaSPValaSPtyrGlInIlePhea 38
   :::::::::::::::::::::
164 GCGACTTGGCGGAAATAAAGCAAGTTTGAGTGGGGGCAAGATAT 213
   :::::::::::::::::::::
38 rgaSPheAlaIaGlInuSnlYsLysPheSerValGIyAlaIthraSnaVal 54
   :::::::::::::::::::::
214 GAGGTNTACAAACAAAAAGGAGTGTGCGCAATCAATGACAAAGC 263
   :::::::::::::::::::::
55 GluValaIraSPlySaAsnaSnaIrgProLeuGIyAsnValaLeuProAsnGI 71
   :::::::::::::::::::::
264 C...CCGATGATTTGATTTTCTGTGGTGTGCG...CGTAACGCGGTGGCGG 307
   :::::::::::::::::::::
71 yIleProMetIleasPheSerValaIaSPValaIaSPlySaIrgIleAlaIat 88
   :::::::::::::::::::::
308 CATGTGTGGCGATCATATATATGTGAGCGTGGCAAT...AAGCGCGGC 354

```

```

88 hrluValaSPnProGIInIryrValaIaGIyAlaIyAlaIySnaIySnaIy 104
   :::::::::::::::::::::
355 TATTAACAACGTTGATTTTGGT.....GCGGAAGCAAGNA 389
   :::::::::::::::::::::
105 ValSerGIuLeuHISpHeGIyAsnLeuAsnGIyAsnMetAsnSngIyAs 121
   :::::::::::::::::::::
390 TCCCGATCAGCAGCGT.....TTTCTTACCAATTTG 421
   :::::::::::::::::::::
121 naIaIySaIaHISaIrgaSPaIySerSerGIuIaSPnaIrgIyrrThy 138
   :::::::::::::::::::::
422 TGAAGAATAATATAT.....AAG 441
   :::::::::::::::::::::
138 alGIuIySaSngIuIryrProThIyLysuAsnGIySaIaValaIthra 154
   :::::::::::::::::::::
442 CTGACATATTCACACCTTACAAAGCGGATTAACATAGCGGCTTGA 491
   :::::::::::::::::::::
155 GluaspGIaIaGIuIySaIrgaIrgIuaspIyrrThyIleProIrgLeuAs 171
   :::::::::::::::::::::
(492 TAAATTTGACAGATGCGACACCAACCTGTCGAAATGACAGATGAC...ATGA 538
   :::::::::::::::::::::
171 pIySPheValThrGIuValaIaProIleGIuAlaSerThraSPSerSert 188
   :::::::::::::::::::::
539 GGGGGAATACCTATTCGATTAAGAAATAATCCGAGCGTGCAGCATC 588
   :::::::::::::::::::::
188 hrlaGIyThryrAsnaSnlYsAspIyrrProIryrPheValaIrgLeu 204
   :::::::::::::::::::::
589 GCGTCAGGACACCATAT.....TGCGCTTATGAT..... 618
   :::::::::::::::::::::
205 GlySerGIyThrGIuIaIleThyGIuAsnGIyThraIrgIyGIuIeUtr 221
   :::::::::::::::::::::
619 .GATGACAAACAGCGGATTTATCTACTCCGCGGCGATG.....TTAA 661
   :::::::::::::::::::::
221 pIeUGIyIySngIyGIuIySnaSnaIrgIyIyThraSnaLeuIySL 238
   :::::::::::::::::::::
662 TTGGCGCAATATACATATATGCAAGGTTGGGAAATATGCGCTATTTACT 711
   :::::::::::::::::::::
238 euValaGIySnaIaIyrr.....ThryrGIyIleAlaGIy 249
   :::::::::::::::::::::
712 TTGAGCGGCGATGTCGCGCATGCCACAGC.....TANGC... 747
   :::::::::::::::::::::
250 ThryrProIryrGIuValaIaHISaIrgIuAsnaSPGIyLeuIleGIyPhe 266
   :::::::::::::::::::::
748 .....CCTA 751
   :::::::::::::::::::::
266 userAsnaSngIuIryrIleAsnProIySngIuIleuSerIySngIySpro 283
   :::::::::::::::::::::
752 TGCCGATTCAGGTGCGGACAGCGGATTCGCCAATGTTATTTAT 801
   :::::::::::::::::::::
283 eutHraSntYrIaValaIleuGIyAspSerGIySerProIeUthraValaIyrr 299
   :::::::::::::::::::::
802 GACAAACAAACAATTAATGCTGCTGCAAGGAGTTTACAAACGGGCA 851
   :::::::::::::::::::::
300 ASPaIrgIuIySngIySnaIrgIuIleuSnaIrgIySnaIrgIySnaIrg 312
   :::::::::::::::::::::
852 CCCTATTCGCGGAGGAAACGGTTTCCAGCTGATACGCAAGAATGGT 901
   :::::::::::::::::::::
312 rAspIyrrTrpIaIaGIyrrAsn.....LysIySerSertPG 324
   :::::::::::::::::::::
902 TCTACATATACATTTACAGAGCGGATACACATACCTGCTCTTTTGAACCG 951
   :::::::::::::::::::::
324 InGIuIryrPnaSnlIeIryrIySngIuIleuAlaGIyIleGIyGIuIa 340
   :::::::::::::::::::::
952 CCAGTAACGGA.....CATTTTCTTCTTACATC 980
   :::::::::::::::::::::
341 TrrSerAlaGIySerLeuIleGIySerIySnaIrgIyrrAspIyrrSerTrp 357
   :::::::::::::::::::::
981 CAACAACAAGGTACGGGTACGTAACAGAAACCAACGAAAGGTTNTCA 1030
   :::::::::::::::::::::
357 rAsnGIy..... 359
   :::::::::::::::::::::
1031 ATCCAAAGTTAAATACAGACAGTCGACTGTTTGACGAATCTTGTGAT 1080
   :::::::::::::::::::::

```

360 .....LysThrSerThrIleThrGlyGlyLysSerLeuAsn 372  
1081 GAAACTGATTAAGAACAGTTCAGCGGAGGGGTGTATTCAGTACCG 1130  
373 .....ValAspLeuAlaAspGly.....Lys 379  
1131 TCAGAGTTAAACAACAGGTTGAAACCTTCTTTATCGATTACGCAACG 1180  
379 sasprysProAsnHisGlyLysSerValThrPhe.....GluGlySerG 394  
1181 GCAACCTATCTTATCAACAACATCAACCAAGCGGCGGCTTGTAT 1230  
394 lYthLeuThrLeuAsnAsnHisLeaspGlnGlyAlaGlyLysLeuPhe 410  
1231 TTGAGGAGATTTTACGTC...TCGCGCTGAAACAAGCAAGAGTGCA 1277  
411 PheGluGlyAspTyrGluValLysGlyThrSerAspAsnThrThrTrrPly 427  
1278 AGCGCGGCGCTTCATATCATGAGACAGTACCGTTACTTGAAGAATA 1327  
427 sGlyAlaGlyValSerValAlaGluGlyLysThrValThrTrrPlyValH 444  
1328 AGCGCGTGGCAACAGCGCGCTGTCCAAATCGGCAAGAGCAGCTGCAC 1377  
444 lAspAsnProGlnTyrAspArgLeuAlaLysIleGlyLysGlyThrLeuIle 460  
1378 GTTCAAGCCAAAGGGGAAACCAAGCTGCATCAGCGTGGGAGAGGTAC 1427  
461 ValGluGlyThrGlyAspAsnLysGlySerLeuLysValGlyAspGlyTh 477  
1428 ACTCATTTTGGATCAGCAGCGCAGAGATTAAGCAAAAAACAGCTTTA 1477  
477 rValIleLeuLysGlnGlnThrAsnGlySerGly...GlnHisAlaPheA 493  
1478 GTGAATCGGCTTGTGTCAGCGCAGCGGTACGTCGAACGTGACCAT 1527  
493 lAspValGlyIleValSerGlyArgSerThrLeuValLeuAsnAspAsp 509  
1528 AATCGATTACCCCGACAACTATTTGGCTTCGCGGCGGACGTTT 1577  
510 LysGlnValAspProAsnSerIleTyrPheGlyPheArgGlyAlaPhe 526  
1578 GGATTTAAACGGGCATTCGTTTCGTCACCGTATTCGAAATACCGATG 1627  
526 uAspLeuAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleAspG 543  
1628 AAGGGCGCATGTGNCNATCATATGCCACACACATCCACGTTTACC 1677  
543 lGluGlyAlaArgLeuValAsnHisSerThrSerLysHisSerThrValThr 559  
1678 ATTACAGGGAATGAAGATTTACACACGAGTGGTAGAATATC... 1722  
560 lIleThrGlyAspAsnLeuIleThrAspProAsnAsnValSerIleTyrTy 576  
1722 ..... 1722  
576 rValLysProLeuGluAspAspAsnProTyrAlaIleArgGlnIleLysT 593  
1722 ..... 1722  
593 yrcGlyTyrGlnLeuTyrPheAsnGlnGluAsnArgThrTyrTyrAlaLeu 609  
1722 ..... 1722  
610 LysLysAspAlaSerIleArgSerGluPheProGlnAsnArgGlyGluSe 626  
1722 ..... 1722  
626 rAsnAsnSerTrrPleuTyrMetGlyThrGluLysAlaAspAlaGlnLysA 643  
1723 .....AATAGACTTAATTACAGCAAGAATAATGCCCTACACGSGTTGG 1764  
643 snAlaMetAsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyr 659

1765 TTGGCGGAGAAAGATGACGACCAAAAGACGGCGGCTCAACCTTGTTA 1814  
660 PheGlyGluGluGlu...GlyLysAsnAsnGlyAsnLeuAsnValThrPh 675  
1815 CCAGCCCGCCGACGAAGACCGCACCCNCGTCTTCCGCGGCAACAAT 1864  
675 eLysGlyLysSerGluGlnAsnArgPheLeuLeuIrrGlyGlyThrAsnL 692  
1865 TAAACGCAACATCACGCAAAACAGCGCAAACTGTTTTCACGCGCAGA 1914  
692 euAsnGlyAspLeuAsnValGlnGlnGlyThrLeuPheLeuSerGlyArg 708  
1915 CCGACACCGCAGCGCTCAATCATTTAGCAAGCGGCTGTCAAAAATGA 1964  
709 ProThrProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysAs 725  
1965 AGGT.....ATCCACACAGGAATAATCGTGTGGACACACGATGATCN 2008  
725 pSerHisPheSerGluAsnAsnGlnValValAlaValAspAspTrrPleA 742  
2009 ACCGCACTTTAAAGCGGAAATTTCCATTTACAGGCGGCGGCGGCTG 2058  
742 snArgAsnPheLysAlaThrAsnIleAsnValThrAsnAlaThrLeu 758  
2059 ATTTCC...CGCAATGTTGCCAAAGTGAAGCGGATTCGATTTGAGCAA 2105  
759 TyrSerGlyArgAsnValGluSerIleThrSerAsnIleThrAlaSerAs 775  
2106 TCACGCCCAAGCAGTTTGGTGTGCGCAGCGCATCAAGCCATACATCT 2155  
775 nAsnAlaLysValHisIleGlyTyr...LysAlaGlyAspThrValC 790  
2156 GTACAGCTGGAGTGCAGNGTCTGACAAATGTGTCGAANAANAAT 2205  
790 yValAlaArgSerAspTrrThrGlyTyrValThrLysStrThrAspLysLeu 806  
2206 ACCGACGATAAGTATGTTGCTTCATTCATGACTAGACAGACNTNAGCGCAN 2255  
807 Ser...AspLysAlaLeuAsnSerPheAsnProThr..... 817  
2256 TGTNAGNCTNCCNATNACNTNNTNNAANCNCNCGGCGNTGCNNCAC 2305  
817 ..... 817  
2306 TNAANGCAATCTTAGTGAATGGCGATACAGTTTACAGTCAGCCAC 2355  
818 .....AsnLeu..... 819  
2356 AACGCCACCCAAAAGCGCAACTTACGCTGTGGGCAATGCCAAGCAAC 2405  
820 .....ArgGlyAsnValAsnLeuThrClnSerIleAsnPheVa 832  
2406 ATTTAATCAACCCACATTAAGCGCAACNATCGGATTCGGGCAATGCTT 2455  
832 lLeuGlyLysAlaAsnLeuPheGlyThrIleGlnSerArgGlyAsnSerG 849  
2456 CATTTAATCTTAGCAACAGCGCGCACAAAAGCGCAGTGTGAGCCTTCC 2505  
849 lValArgLeuThrGluAsn..... 855  
2506 GACAACGCTAAGCAAAAGTAAAGCATTCGCGACTCAAGGCAATGTCTC 2555  
855 ..... 855  
2556 CCTAGCCGATAGGACAGATTCCATTTGAAACAGCCGCTTTACCGGAC 2605  
855 ..... 855  
2606 AACTCAGCGCAGCAGAGANACAGCATTTACACTTAAAGACAGGAGATG 2655  
856 .....SerHisTrr 858





```

1422 ILehrProIIeValGlyValArgTyrSerTyrLeuSerAsnAlaaspPh 1438
4020 CCGGTACGAAACGTCATATCGCCACCCGGCTGCTGCTTCAACCGNT 4069
      : : : : : : : : : : : : : : : : : : : : : : : :
1438 eAlaLeuaspGlnAlaArgIleLysValAsnProIIeSerValIlysthrA 1455
4070 ACCGNGCGGCAATTAAGCAGATTAATTCATTCMAACCGGCCAACACATN 4119
      : : : : : : : : : : : : : : : : : : : : : : : :
1455 IapheAlaGlnValAsnLeuSerTyrThrTyrHisLeuGlyGlu...Phe 1470
4120 TCCATACACNCCTTATTAGCCTGCTCATACCGATGCCGCTGGCGCAA 4169
      : : : : : : : : : : : : : : : : : : : : : : : :
1471 SerValIhrProIIeLeuSerAlaArgTyr...AspAlaAsnGlnGlySe 1486
4170 AGTCGACAGCGGCTCATACCGGCTATGTCGATTCAGATTCGGCAAAA 4219
      : : : : : : : : : : : : : : : : : : : : : : : :
1486 IGLyLysIleAsnValAsnGlyTyrAspPheAlaTyrAsnValGluAsnG 1503
4220 CCGCGAGTGGCGAATGAGCGGTAAACGCCGAATCAAGCTTTCACGCTG 4269
      : : : : : : : : : : : : : : : : : : : : : : : :
1503 InGlnGlnTyrAsnAlaGlyLeuLysLeuTyrHisAsnValLysLeu 1519
4270 TCCNTCCACGCTGCCGCCGCCCAAGNCCGCACTGCAAGCGCAACAG 4319
      : : : : : : : : : : : : : : : : : : : : : : : :
1520 SerLeuIleGlyGlyLeuThrLysAlaLysGlnAlaGlnLysGlnLysTh 1536
4320 CGCGGCGATCAATTAAGGCTAC 4341
      : : : : : : : : : : : : : : : : : : : : : : : :
1536 rAlaGlnLeuLysLeuSerPhe 1543

seq_name: SwissProt_40: IGAL_HAEIN

seq_documentation_block:
ID IGAL_HAEIN STANDARD; PRT; 1541 AA.
AC PA2782;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-HK368 / SEROTYPE B;
RX MEDLINE=89379374; PubMed=2506130;
RA Poulsen K., Brandt J., Hjorth J.P., Thøgersen H.C., Kilian M.;
RT "Cloning and sequencing of the immunoglobulin A1 protease gene (iga)
RT of Haemophilus influenzae serotype b.";
RL Infect. Immun. 57:3097-3105(1989).
RN [2]
RP MUTAGENESIS OF SER-288.
RX STRAIN-HK368 / SEROTYPE B;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases.";
RL J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; X64357; CAAB5708.1; -
DR EMBL; M87492; AAA24969.1; -
DR MEROPS; S06.001; -
DR InterPro; IPR000710; IGA_S6.
DR Pfam; PF02395; IGAL; 1
DR PRINTS; PR00921; IGASERPTASE.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1008
FT PROPEP 1009 1541
FT ACT_SITE 288 288
FT MUTAGEN 288 288
FT SEQUENCE 1541 AA; 169370 MW; CE7257CB3196C600 CRC64;

```

```

alignment_scores:
Quality: 1536.50 Length: 1722
Ratio: 1.677 Gaps: 56
Percent Similarity: 53.194 Percent Identity: 26.887

```

```

alignment_block:
US-09-303-518D-651 x IGAL_HAEIN

```

```

Align seg 1/1 to: IGAL_HAEIN from: 1 to: 1541

```

```

64 CGCTTCTCGCCCTGCTACTTACGCAATATGCTGCTGCGCATTCCTCC 113
      : : : : : : : : : : : : : : : : : : : : : : : :
5 LysPheLysLeuAsnPhelIleAlaLeuThrValAlaTyrAlaLeuThrPr 21
114 CCAAGCTGGCGGACACACTATTCGGCATCACTACCAATACTATAC 163
      : : : : : : : : : : : : : : : : : : : : : : : :
21 oTyrThrGlnAlaAlaLeuValAlaArgAspAspValAspTyrGlnIlePheA 38
164 GCGACTTGGCGGAATTAAGCAAGTTCGAGTCGGCGCAAGATATT 213
      : : : : : : : : : : : : : : : : : : : : : : : :
38 rGaspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54
214 GAGCTTACACAAAAGAGGAGTGTGCGCAATATGACAAAGC 263
      : : : : : : : : : : : : : : : : : : : : : : : :
55 LeuValLysAspLysAsnAsnLysAspLeuGlyThrAlaLeuProAsnG1 71
264 C...CCGATGATGATTTTCTGCTGCTGCG...CGTACGGCGTGGCGCG 307
      : : : : : : : : : : : : : : : : : : : : : : : :
71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88
308 CATTTGGTGGCGCATATATTTGTGAGCGGACAT...AACGGCGCG 354
      : : : : : : : : : : : : : : : : : : : : : : : :
88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104
355 TATAACAAGCTGATTTTGTG...GCGGAAGGAAGNNA 389
      : : : : : : : : : : : : : : : : : : : : : : : :
105 ValSerGlnLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121
390 TCCCGCATGACACCGT...TTTCTTACCAAAATG 421
      : : : : : : : : : : : : : : : : : : : : : : : :
121 nAlaLysAlaHisArgAspValSerSerIleGlnAsnAlaGlyTyrPheSerV 138
422 TGAAGAAGAAATAATAT...AAG 441
      : : : : : : : : : : : : : : : : : : : : : : : :
138 aGlnLysAsnGlnLysTyrProThrLysLeuAsnGlyLysThrValThrThr 154
442 CCGCAATATTCACACCTTACAAAGCGGCTTAMCATATATGCCGCTTGCA 491
      : : : : : : : : : : : : : : : : : : : : : : : :
155 GluAspGlnThrGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171

```

```
492 TAAATTTGTCAGAGATGCAAGAACTGTGCA...ATGACGAGTCAGCATGA 538
|||||
171 pypserValThrGluValAlaProIleGluIleSerThrAlaSerSerA 188
|||||
539 GGGGAAATCCATTCGCCATTAAGAAAAATATCCGACGCTCCGCAATC 588
|||||
188 spIleGlyThrIleAsnAspGlnAsnIleGlySerProAlaPheValArgLeu 204
|||||
589 GCGTCAGACACCACTATTGGCGTTATGATGATGACAAACAGCGGATTT 638
|||||
205 GlySerGlySerGlnPheIleTyr.....LysLysGlyAsp... 216
|||||
639 ATCCATCTCCGGCGCATGGTTAATTGGCGGCAATACATATGACAGGTT 688
|||||
217 .AsnYrSer.....LeuIleLeuAsnAsnIleGluValGly. 228
|||||
689 GGGGAAATTAAT.....GGCGTAAAT 708
|||||
229 ..GlyAsnAsnLeuLysLeuValGlyAspAlaTyrThrTyrGlyIleAla 244
|||||
709 AGTTTGAGCGCGCATGTCCGCGCATGCCAACGAC.....TATGG 746
|||||
245 GlyThrProTyrLysValAsnIleGluAsnAsnGlyLeuIleGlyPheGly 261
|||||
747 C.....C 748
|
261 yAsnSerLysGluGlnIleHisSerAspProLysGlyIleLeuSerGlnAsp 278
|
749 CTATGCGCGATTCGAGGTGGCGGACGCGCATGCCCAATGTTTAT 798
|||||
278 roLeuThrAsnTyrAlaValIleuGlyAspSerGlySerProLeuPheVal 294
|||||
799 TATGACAAACAAACAATMAATGCTGCTCAACGAGATTTCACAAACGG 848
|||||
295 TyrAspArgGlnLysGlyLysTrpLeuPheLeuGly.....Se 307
|||||
849 CTACCTTATTCGCGACGAGGAAACGGTTTCCAGCTGATACGCAAGATT 898
|||||
307 rTyrAspPheTrpAlaGlyTyrAsn.....LysLysSer 319
|||||
899 GGTTCACGATGACATTTACAGAGCGCATACATACCGTCTNTTTGAA 948
|||||
319 rGlnGlnIleTrpAsnIleTyrLysSer..... 327
|||||
949 CGCGCGACGTAAACGACATTTTCCTTTACATCC.....ACACAGAA 989
|||||
328 .....GlnPheThrLysAspValIleAsnLysAs 337
|||||
990 CGGTACGGGTACGGTACAGAAACCAAGAAAGTATCC...AATCCA 1036
|||||
337 pserIleGlySerLeuIleGlySerLysThrAspTyrSerTrpSerSerA 354
|||||
1037 ACCTTAAGTACAGACAGTCCGACTGTTTGAAGATCTTTGAATAAAT 1086
|||||
354 snGlyLysThrSerThrIleThrGlyGlyGluLysSerLeuAsn..... 368
|||||
1087 GATAAAGAACAGTTTACCGCGGACGGGGTGTAAATCAGTACCGTCAAG 1136
|||||
369 .....ValAspLeuAlaAspGly.....LysAspLys 377
|||||
1137 GTTAAACAACGGTGAACCTTTCTTATGATGATACGCAACGGCAAC 1186
|||||
377 spIAsnHisGlyLysSerValThrPhe.....GlnGlySerGlyThrL 392
|||||
1187 TCATGTTTATCAACAACATCAACCAAGGCGGCGGTTGTGATTTGAA 1236
|||||
392 euThrLeuAsnAsnAsnIleAspGlnGlyAlaGlyGlyLeuPhePheLeu 408
|||||
1237 GGTGATTTTACGGTC...TCGCGTGAACAAACGAACGTGCGCAAGCGC 1283
|||||
409 GlyAspTyrGluValLysGlyThrSerAspAsnThrThrTrpLysGlyAl 425
|||||
1284 GGGCGTTCAATATCACTGAAGACGTACCGTTACTTGGAAGTAAACGGCG 1333
|||||
425 aglyValSerValAlaGlnIleLysThrValThrTrpLysValHisAsnPr 442
|||||
1334 TGGCAACAGCCCGCTGTCCAAAATCGGCAAGGACGCGTCAGCTTCAA 1383
|||||
442 roGlnTyrAspArgLeuAlaLysIleGlyLysGlyThrLeuIleValGlu 458
|||||
1384 GCCAAAGGGGAAAAACCAAGGCTGCATCGCTGGCGGACGCTGATCAT 1433
|||||
459 GlyThrGlyAspAsnLysGlySerLeuLysValGlyAspLysThrValI 475
|||||
1434 TTTGGATCAGCAGGACAGCATTAAGGCAAAAAACAAGCTTATAGTAAA 1483
|||||
475 eLeuLysGlnGlnThrAsnGlySerGly...GlnHisAlaPheAlaSerY 491
|||||
1484 TCGCGTTGNTCAGCGGCGCATCGGTACGCTGCACTGCAATGCCATTAAC 1533
|||||
491 alGlyIleValSerGlyArgSerThrLeuValLeuAsnAspAspLysGln 507
|||||
1534 TTCACCCCGCAACAACCTATTCGCGCTTTCGCGGCGGACGTTGGATT 1583
|||||
508 ValAspProAsnSerIleTyrPheGlyPheArgGlyGlyArgLeuAspLe 524
|||||
1584 AACGGGCACTTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGG 1633
|||||
524 uAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleAspAspLysA 541
|||||
1634 CGATGATTGCMNATCATATGCCAACACATCCACCGTTACCATTTACA 1683
|||||
541 laArgLeuValAsnHisAsnMetThrAsnAlaSerAsnIleThrIleThr 557
|||||
1684 GGGATGAAGATTTACACACCGAGT..... 1710
|||||
558 GlyLysSerLeuIleThrAspProAsnThrIleThrProTyrAsnIleAs 574
|||||
1710 ..... 1710
|||||
574 palAProAspGluAspAsnProTyrAlaPheArgArgIleLysAspGlyG 591
|||||
1710 ..... 1710
|||||
591 LysGlnLeuTyrLeuAsnLeuGlnAsnTyrThrTyrTyrAlaLeuArgLys 607
|||||
1710 ..... 1710
|||||
608 GlyAlaSerThrArgSerGlyLeuProLysAsnSerGlyGluSerAsnGly 624
|||||
1711 .....GTTAAGAT.....A 1720
|||||
624 uAsnTrpLeuTyrMetGlyLysThrSerAspGluAlaLysArgAsnValH 641
|||||
1721 TCATAGACTTAATTTACAGCAAGAAATTCCTACACAGGTTGGTTGGC 1770
|||||
641 etAsnHisIleAsnAsnGlnIleArgMetAsnGlyPheAsnGlyTyrPheGly 657
|||||
1771 GAGAAAGATACGACCAACCAAGGCGGCTCAACCTGTTTACAGGCC 1820
|||||
658 GlnGlnGln...GlyLysAsnAsnGlyAsnLeuAsnValThrPheLysGly 673
|||||
1821 CGCGCGACAGACCGGACCCGCTGCTTCCGCGGACCAACAAATTTAAAG 1870
|||||
673 LysLysSerGlnGlnAsnArgPheLeuLeuThrGlyGlyThrAsnLeuAsnG 690
|||||
1871 GCAACATCATCAACAAACAGGCAAACTGTTTTCAGCGGACAGCGACAGCA 1920
|||||
690 LysAspLeuThrValGlnLysGlyThrLeuPheLeuSerGlyArgProThr 706
|||||
1921 CCGCAGCGCTTACATCATTTAGAGAGCGGCTGCTCAAAAATGGAAGCT.. 1968
|||||
707 ProHisIleArgAspIleAlaGlyIleSerSerThrThrLysLysAspProH 723
|||||
1969 ....ATCCCAAGGAGAAATGCTGGGCAACGACTGGATGNACCGCA 2014
|||||
```

```
723 sPheLaGluAsnAsnGluValValGluAspSPTrPLeAsnArgA 740
2015 CGTTTAAACGGAAATTCATTCATTCAGGGCGGCGGTGATTCC 2064
740 snPheLysAlaThrMetAsnValThrGlyAsnAlaSerLeuTyrSer 756
2065 ...CGCAATGTTGCCAAAGTGGAGGCGATTCGATTGAGCAATCAGC 2111
757 GlyAlaGlnValAlaAsnIleThrSerAsnIleThrAlaSerAsnIle 773
2112 CGAAGCAGTTTTGTGTGCGACCCGATCAAGCCATCAATCTGTAC 2161
773 agIlnValHisIleGlyTyr....LysThrGlyAspThrValGlyVal 788
2162 GTTGGCAGTGGACNGCTGCAAAATGTGTGCAAAANCAATTTbCCG 2211
788 rGserAspTyrThrGlyTyrAlaThrCysThrThrAspLysLeuSer... 803
2212 GATTAAGTATGCTTCATTCATTAAGACNGACNTNAGCGCANTGTNAG 2261
804 AspLysAlaLeuAsnSerPheAsnProThr..... 813
2262 NCTNNCCNATNACGNTNNTTNAANCNCNGGCGNTGCGNCACTNANG 2311
813 ..... 813
2312 GCAATCTTAGTCAATGCGCATACAGCTTATACAGTACGCCACAAGCC 2361
814 ...AsnLeu..... 815
2362 ACCCAAAACGCAACCTTAGCTCGTGGGCAATGCCCAACAACATTTAA 2411
816 ....ArgGlyAsnValAsnLeuThrGluSerAlaAsnPheValLeuGI 830
2412 TCAAGCCCATTTAAACGCAACNCATCGGNTTGGGCAATGCTTCATTTA 2461
830 yLysAlaAsnLeuPheGlyThrIleGlnSerArgLysAsnSerGlnValA 847
2462 ATCTAAGCAACAACGGCGGCAAAACGGAGTCTGACGCTTCCGACAAC 2511
847 rGleuThrGluAsn..... 851
2512 GCTAAGCAAAAGTAAGCCATTCGCGACTCAAGGCAATGTCTCCCTAGC 2561
851 ..... 851
2562 CGATTAAGCAATATTCATTTTGAACAACGCGCTTTACGGACAATCA 2611
851 ..... 851
2612 GCGGCGCAAGCAAGNACGATTACACTTAAAGACGCAATGACGCTG 2661
852 .....SerHisThrPheLeu 856
2662 CGGTGACGACGGAATTTAGCAATTTAACTTGACAACGCCACCATTTAC 2711
857 ThrGlyAsnSerAspValHisGlnLeuAspLeuAlaAsnGlyHisIleH 873
2712 ACTCAATTCGCGCTATCGCCACGATGCTGACGGCGGCAACCGGCAANG 2761
873 sLeuAsnSerAlaAspAsnSerAsn..... 881
2762 TGTCNACACAGCGCGCGCGCGCTTCCTATTATCCGTTACA 2811
882 .....AsnValThr 884
2812 CGGCCAATTCGTTAGTAATCCGTTTCAACAGCTGACGTAAGGCA 2861
885 .....LysTyrAsnThrLeuThrValAsn...Se 893
2862 ATTGAACNGTCAAGACAATTCGCTTATGTGGAACCTCTTGCGCTACC 2911
893 rLeuSerGlyAsnGlySerPheTyrThrLeuThrAspLeuSerAsnLysG 910

2912 GAAGCAAAATTTAGACTGGCGGAAAGTTCCGAGGNACTTACACCTTG 2961
910 InGlyAspLysValValValThrLysSerAlaThrGlyAsnPheThrLeu 926
2962 GCGGTCAACATATACCGGCAACGACCCGTAAGCCCTCATCAATTTGACG 3011
927 GlnValAlaAspLysThrGlyGluPro...AsnHisAsnIleuThrLe 942
3012 AGTGAAGGGAAGACAACAACCCGTCGCGCAAAACCTTAATTTACCC 3061
942 uPheAspAla....SerLysAlaGlnArgAspHisLeuAsnValSerL 957
3062 TGCAAAACGAAACACAGTCCGCGCGGCGCTGCGCTTCCAACTCATCCG 3111
957 euValGlyAsnThrValAspLeuGlyAlaThrPlyTyrLysLeuArgAs 973
3112 AAGACGCGCAGTTCGCGCTGCATTAATCCGTCMAAGACA..... 3153
974 ValAsnGlyArgTyrAspLeuTyrAsnProGluValGluLysArgAsnGI 990
3153 ..... 3153
990 nThrValAspThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspV 1007
3154 .....GAGCTTCCGACAACACTCGGCAAG..... 3177
1007 alProSerValProSerAsnAsnGluIleAlaArgValAspGluAla 1023
3177 ..... 3177
1024 ProValProProProAlaProAlaThrProSerGluThrThrGluThrVa 1040
3178 .GCAGAACCAAAAACAGCGCGA.....GACACCGCGCAA 3217
1040 lAlaGluAsnSerLysGlnSerLysThrValGluLysAsnGluIlnA 1057
3218 GCGTGAACGCGCTGATTCGCGCGCGGCAATGCGCGCGAA..... 3258
1057 sPalaThrGluThrAlaGlnAsnArgGluValAlaLysGluAlaLys 1073
3259 .....AAGACAGAAAGCGTTCGCGAACCAGCGCGCG 3287
1074 SerAsnValLysAlaAsnThrGlnThrAsnGluValAlaGln..... 1087
3288 GCGGACGCGCGGGAATGTCGCAATTATGACGCGGAG..... 3327
1088 ...SerGlySerGluThrLysGluThrGlnThrThrGluThrLysGluT 1103
3328 .....GAAGCAAAAACGCGTTCGCGAACCAGCGCGG 3360
1103 hrAlaThrValGluLysGluIleLysValLysValGluThrGluLysThr 1119
3361 AGCGCNTTGGCGAAA.....CAGCGGCAACCGGGAACCCGCGCGN 3401
1120 GlnGluValProLysValThrSerGlnValSerProLysGlnGluInse 1136
3402 TACACGCGCTTCCCGCGCGCGCGCGCGCGCGGGAATTTGCGG.... 3447
1136 rGluThrValGlnProGlnAlaGluProAlaArgGluAsnAspProThrV 1153
3448 .....CAACGCGACGCCCA..... 3462
1153 alaAsnIleLysGluProGlnSerGlnThrAsnThrThrAlaAspThrGlu 1169
3463 .....CCGCAACT..... 3471
1170 GlnProAlaLysGluThrSerSerAsnValGluInProValThrGluSe 1186
3471 ..... 3471
1186 rThrThrValAsnThrGlyAsnSerValValGluAsnProGluAsnThrT 1203
```

```

3472 .....CAACC.....CAACG 3483
1203 hrProAlaThrThrGlnProThrValAsnSerGluSerSerAsnLysPro 1219
3484 CAGGC..... 3489
1220 LysAsnArgHisArgSerValArgSerValProHisAsnValGluPr 1236
3490 .....GACCTGA 3496
1236 oAlaThrThrSerSerAsnAspArgSerThrValAlaLeuGlyAspLeuT 1253
3497 TNAAGCGTTATGCCAATAGCGGTTGATGATTTTCCGCC..... 3537
1253 hrSerThrAsnThrAsnAlaValLeuSerAspAlaArgAlaLysAlaGln 1269
3538 .....ACGCTACACAGCGTTTCCGCTACAGAGACGAAATGGACCGCT 3581
1270 PheValAlaLeuAsnValGlyLysAlaValSerGlnHisIleSerGlnLe 1286
3582 G...TTTGGCAGACCGCGCAACGCGTTTGGACACGCGCATCCGCGN 3628
1286 uGluMetAsnAsnGluGlyGlnIlyrAsnValIlyrValSerThrSerM 1303
3629 ACACCAACACACTACCGCTGCAAGATTTCCGCGCTACCGCAACCAAC 3678
1303 etAsnLysAsnIlyrSerSerGlnIlyrArgArgPheSerIlyrSer 1319
3679 GACCTGGCGCAATCGCTATGCAAAAAACCTCGGACGCGG...CGCCT 3725
1320 ThrGlnThrGlnLeuGlyIlyrAspGlnIlyrIleSerAsnAsnValGlnLe 1336
3726 CGGCATCTGTTTTCGCAACACGCGACGCAAAACNTTCGACAGCGCA 3775
1336 uGlyValAlaPheThrIlyrValArgAsnSerAsnAsnPheAspLysAlT 1353
3776 TCGGCACTCGGACGCGCTTCCCGCGCGCGCTTTCGGCAATACGCG 3825
1353 hrSerLysAsnThrLeuAlaGlnValAsnPheIlyrSerIlyrIlyrAla 1369
3826 ATGGCGAGGTGCGACATCGGCATACACGCGCGCGCTTATAGCGCGG 3875
1370 AspAsnHisTrpIlyrLeuGlyIleAspLeuGlyIlyrGlyLysPheGlnLe 1386
3876 CAATCTNTGAGACGCGCATCGGACGCAAAATCCGCGCGCGCTGCTCAT 3925
1386 IlyrLeuGlnIlyrAsnHisAsnAlaLysPheAlaArgHisIlyrAlaGlnP 1403
3926 ACGGCATTACGCGCATACCGCGCGCTTTCGGCGGATTCGCGCATCGCA 3975
1403 heGlyLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnPheGlyIlyrThr 1419
3976 CGGTACATGCGCGCGCAACGCGCTATTTCGCAAAAAACGCGATTACCGCA 4025
1420 ProIleValAlaGlyAlaGlyIlyrSerIlyrLeuSerAsnAlaAspPheAlaLe 1436
4026 CGAAAAAGCTAATATGCGCACCGCGCTTTCGCTCAACGCTACCGCG 4075
1436 uAspGlnAlaArgIleLysValAsnProIleSerValLysThrAlaPheA 1453
4076 CGGCGATTAGGCGAGATTATTCATTTCACACGCGCGCAACACATNTCCATC 4125
1453 IleGlnAlaLysPheSerIlyrIlyrHisLeuGlyGlu...PheSerVal 1468
4126 ACNCGCTATTNAGCGCTGCTACGCGATACCGGCTTTCGGCGCAAGTCCG 4175
1469 ThrProIleLeuSerAlaArgIlyr...AspAlaAsnGlnGlySerGlyLys 1484
4176 AACACGCGCTAATACGCGNGTATGCTAGCATGTTTTCGCAAAACCGCGA 4225
1484 sIleAsnValAsnGlyIlyrAspPheAlaIlyrAsnValGluAsnGlnGlnG 1501
4226 GTGCGGATGCGGCGTAAACGCGCAAAATCAAGGTTTTCACGCTGTCNCTC 4275

```

```

1501 IntyAsnAlaGlyLeuLysLeuLysTrpHisAsnValLysLeuSerLeu 1517
4276 CACGCTGCGCGCGCGCAAGNCCGCACTGGAACGCGCAACACACGCGCGG 4325
1518 IleGlyGlyLeuThrLysAlaLysGlnAlaGlnLysGlnLysThrAlaG 1534
4326 CATCAATATTAGGCTAC 4341
1534 uLeuLysLeuSerPhe 1539

seq_name: SwissProt_40:IGA0_HAEIN
seq_documentation_block:
ID IGA0_HAEIN STANDARD; PRT; 1694 AA.
AC P44969;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA OR IGA1 OR HI0990.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEROTYPE D;
RA Wright A., Fishman Y., Tai F., Plaut A.G.;
RL Submitted (MAY-1991) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodok A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ufferback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fritchman J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; X59800; ?; NOT ANNOTATED_CDS.
CC EMBL; U32779; AAC22651.1; -
CC MEROPS; S06.001; -
CC TIGR; HI0990; -
CC InterPro; IPR000710; IGA_S6.
CC Pfam; PF02395; IGA1; 1.

```

DR PRINTS; PRO0921; IGASERPTASE.  
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal

FT	SIGNAL	1	25	POTENTIAL IMMUNOGENIC AL PROTEASE HELPER PEPTIDE (POTENTIAL) PROBABLE.
FT	CHAIN	26	1014	G -> A (IN REF. 1).
FT	PROPEP	1015	1694	G -> E (IN REF. 1).
FT	ACT_STATE	288	288	S -> T (IN REF. 1).
FT	CONFLICT	253	254	A -> D (IN REF. 1).
FT	CONFLICT	272	272	A -> D (IN REF. 1).
FT	CONFLICT	464	464	A -> G (IN REF. 1).
FT	CONFLICT	866	866	A -> T (IN REF. 1).
FT	CONFLICT	1036	1036	A -> D (IN REF. 1).
FT	CONFLICT	1074	1074	A -> G (IN REF. 1).
FT	CONFLICT	1421	1421	A -> T (IN REF. 1).
FT	CONFLICT	1545	1545	H -> Q (IN REF. 1).
SO	SEQUENCE	1694 AA:	185539 MM:	C52A2427013F53178C CR664

```
alignment_scores:      Length: 1873
    Quality: 1519.00
    Ratio: 1.653       Gaps: 54
Percent Similarity: 49.066 Percent Identity: 24.826
```

```
alignment_block:
```

US-09-303-518D-651 x IGA0\_HAEIN .

Align seg 1/1 to: IGA0\_HAEIN from: 1 to: 1694

[illegible][illegible]

441 hrTrpIysValHisAsnProLysSerAspArgLeuAlaLysIleGlyLys 457  
1366 GGCACGCTCAGCTTCACAGCAAGGGGAAACCAAGGCTGCATCAGCGT 1415  
458 GlyThrLeuIleValGluGlyLysGluAsnLysGlySerLeuLys 474  
1416 GGGCGACGGTAGATCATTTTGGATCAGCAGCAGCAGATAAAGCAAAA 1465  
474 IgIAspGlyThrValIleLeuLysGlnGlnIleAspAlaAsnAsnLysV 491  
1466 AACAAACCTTTAGTGAATCGCTGNTCAGCGCAGGCGTACGGTCCAA 1515  
491 aLysAlaSerGlnValGlyIleValSerGlyArgSerThrValVal 507  
1516 CTGAATGCGCATATCATCTTCAACCCGCAACCTCTATTTGGCTTTCG 1565  
508 LeuAsnAspAspLysGlnValAlaAspProAsnSerIleTyrPheGlyPheArg 524  
1566 CGGCGGACGTTTGGATTTAAAGGGCATTGCTTTCGTTCCACCGTATTC 1615  
524 gGlyGlyArgLeuAspAlaAsnGlyAsnAsnLeuThrPheGluHisIleA 541  
1616 AAAATACCGATGAAAGGGCGATGATTCATCATTAATCCCAACAACA 1665  
541 rGAsnIleAspAspGlyAlaArgLeuValAsnHisAsnThrSerLysThr 557  
1666 TCCACCGTATTCATTCAGGGAATGAATATTACACACGAGT..... 1710  
558 SerThrValThrIleThrGlyLysSerLeuIleThrAspProAsnThrII 574  
1710 ..... 1710  
574 eThrProTyrAsnIleAspAlaProAspGluAsnProTyrAlaPheA 591  
1710 ..... 1710  
591 rGArgIleLysAspGlyGlyGlnLeuTyrLeuAsnLeuGluAsnTyrThr 607  
1710 ..... 1710  
608 TyrTyrAlaLeuArgLysGlyAlaSerThrArgSerGluLeuProLysAs 624  
1711 .....GTTAAGAAAT..... 1719  
624 nSerGlyLysSerAsnGluAsnTyrLeuTyrMetGlyLysThrSerAspG 641  
1720 .....ATCAATAGACTTAATTACAGCAAGAATAATGCC 1752  
641 LuAlaLysArgAsnValMetAsnHisIleAsnAsnGluArgMetAsnGly 657  
1753 TACAACGGTTGGTTGGCAGAAAGATAGCAACCAAAAGGAGCGCGCT 1802  
658 PheAsnGlyTyrPheGlyGluGluGlu...GlyLysAsnAsnGlyAsnLe 673  
1803 CAACCTGTTTACCGCCCGCGCAGAGACCCGACCCNGCTGCTTCG 1852  
673 uAsnValThrPheLysGlyLysSerGluGlnAsnArgPheLeuLeuThrG 690  
1853 GCGGAACAAATTTAAACGCAACATCAGCAACAAACAAACGCAAACTGTT 1902  
690 LyGlyThrAsnLeuAsnGlyLysPheLysValGluLysGlyThrLeuPhe 706  
1903 TTCAGCGGACAGCAGCAGCGCGCTTACATATTTAGGAAGCGGTG 1952  
707 LeuSerGlyArgProThrProHisAlaArgAspIleAlaGlyIleSer 723  
1953 GTCAAAAATGAGAGT.....ATCCCAACAGCAAAATCGTGGGACA 1996  
723 rThrLysLysAspGlnHisPheAlaGluAsnAsnGluValValGluVal 740  
1997 ACAGCTGATNACCGCAGCTTTAAAGCGAAATTTTCATATTCAGGCG 2046  
740 spAspTrpIleAsnArgAsnPheLysAlaThrAsnIleAsnValThrAsn 756  
2047 GGCAGCGCGGTGATTTCC...CGCATCTTCCCAAGTGCAGAGCGCATTC 2093  
757 AsnAlaThrLeuTyrSerGlyArgAsnValAlaAsnIleThrSerAsnII 773  
2094 NCATTTGAGCATATCAGCCCAAGCAGATTTTGGTGGTGCACCGCATCAAA 2143  
773 eThrAlaSerAspAsnAlaLysValHisIleGlyTyr.....LysAlaG 788  
2144 GGCATACATCTGTACAGCTTTCGACTGACAGGCTGCACAAATTTGTC 2193  
788 LysThrValLysValArgSerAspTyrThrGlyTyrValThrCysThr 804  
2194 GAANAANAACATTCACGACGATTAAGTATGCTTCATGACTAGACNCA 2243  
805 ThrAspLysLeuSer...AspLysAlaLeuAsnSerPheAsnAlaThrAs 820  
2244 CTTNAGCGGCGCATNTNAGCTNNCCNATNAGTNNNTNAAACCTCNCNG 2293  
820 nValSer..... 822  
2294 GGCNTGCNNCATNANGCAATCTTAGTGAATAGCGATACAGCTTAT 2343  
822 ..... 822  
2344 ACAGTACAGCCAAACGCCCAACCAACCGCAACCTTAGCTGTGGGCA 2393  
823 .....GlyAsnValAsnLeuSerIleLys 830  
2394 TGCCCAAGACACATTTAATCAAGCCACATTTAAAGCGCAACNCATCGGNT 2443  
830 nAlaAsnPheValLeuGlyLysAlaAsnLeuPheGlyThrIleSerGlyT 847  
2444 CGGCAATGCTTCATTTAATCTAAGCAACACACCGCCGACAAAAGCGACT 2493  
847 hrGlyAsnSerGlnValArgLeuThrGluAsn..... 857  
2494 CTGACGCTTTCGCAACAGCTTAAGCAACGTAAAGCATTCGCACTCAA 2543  
857 ..... 857  
2544 CGGCATGTCTCCCTAGCCGATAGGAGCATTTTCATTTGAAACAGCC 2593  
857 ..... 857  
2594 GCTTTACCGGACAACTCAGCGGACAGCAAGSANNACGATTACCTTAAA 2643  
857 ..... 857  
2644 GACAGGAATGAGACGCTGCCGTACGACGAGCAATTAAGCAATTTAAACCT 2693  
858 ...SerHisTrpHisLeuThrGlyAspSerAsnValAsnGlnLeuAsnLe 873  
2694 TGAACAACGCCCATTTACACTCAATTCGCGCTATTCGCCACAGATGCTGAC 2743  
873 uAspLysGlyHisIleHisLeuAsnAla.....GlnAsnAspAlaAsnL 888  
2744 GCGCGCAACCGCGCAGNGTGTGACACAGCGCGCGCGCTTCGCGCGCT 2793  
888 yValThrThr.....TyrAsnTh 891  
2794 TCCCTATTATTCGTTACACCGCCAACTTCGTAGATCCCGTTTCACAC 2843  
892 ..... 894  
2844 GCTGACGGTAAAGCAATTTGACNGTCAAGSACATTCGCTTTATGT 2893  
894 rLeuThrValAsn...SerLeuSerGlyAsnGlySerPheTyrTyrLeuT 910  
2894 CGGAACCTCTTCGGCTACGAGGAGCAAAATTTAGAGTGGCGGAAGTTCC 2943  
910 hrAspLeuSerAsnLysGlnGlyAspLysValValValThrLysSerAla 926





```

1523 ASPASNHTPTPTLYLEUQIYLLEASPLEUQIYTYGILYSPHEGLNSE 1539
3876 CANTTCNTFCAGAGCATCGAGGCAAAATCCGCCCGCGTCGCATTC 3925
1539 IANLEULYSTPTASNHISANLALYSAPHEALATGHSITPILAGLNP 1556
3926 ACCGCAATTCAGGACAGATACCGCGCGTTTCGGCGGATTCGCGATCGAA 3975
1556 hEILYUHTHRLAGILYLSALAPHEANLEUQIYASNPHGLYLETHT 1572
3976 CGGTATACGCGGCAACGGCGTATTCGTCGCAAAAAGCGGATACCGGTA 4025
1573 PTLIEVALIGLYLALATGTYSERTYTLTYSERASNLALASNPHEALALE 1589
4026 CGAAACGTCATATCGCCACCCCGCTTCGTCGTCACCCNTAACCGNG 4075
1589 uALALYSAPRATGILIELYVALSNPPTLIESERVALYSTHRLAPHEA 1606
4076 CGGCGATTAGGCGATATTCATTCAAAACGCGGCAACATNTCCATC 4125
1606 lAGLVALASPLEUSERTYTLTHTYHISLEUGILYU...PHESEVAL 1621
4126 ACGCTATTTCAGCCTGTCTAT...ACCGATGCCGCTTCGGCGCAAGT 4172
1622 THPTLIEUSERALATGTYSRPTASNGLNGLYSERGLYLSIL 1638
4173 CGGACACGCGTCATATCCGCGTATTCGTCAGATTCGCGCAAAACCC 4222
1638 e.....ASNVALASNGLNTYASRPHALATYASNPVALIGLVALSNGING 1653
4223 GCAGTCCGGAATGGCGGTAAACGCCGAATCAAGATTACAGCTGTCC 4272
1653 lNGLNTYASNPALAGILYUHTYLSLEYSTYHISASNPVALYLSLEUSER 1669
4273 NTCCACGCTGCCGCGCGCAAGNCCGCAATCGAAAGCCAGCACACAGCGC 4322
1670 LEULIEGLIGLYUHTYLSALYSGINALAGILYLSGLNLYSTHRL 1686
4323 GGGCATCAATTAAGGCTAC 4341
1686 aGLIULEULYSLEUSERPHE 1692
seq_name: SwissProt.40:IGA2_HAEIN
seq_documentation_block:
ID IGA2_HAEIN STANDARD; PRT; 1702 AA.
AC P45384;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=127;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK715 / SEROTYPE B;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
influenzae type 1 immunoglobulin A1 proteases."
RL J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
certain Pro-I-Xaa bonds in the hinge region. No small molecule
substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE

```

```

CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M87489; AAA24966.1; -.
CC DR MEROPS; S06.001; -.
CC DR InterPro: IPR000710; IGA_S6.
CC DR Pfam: PF02395; IGA1; 1.
CC DR PRINTS; PRO0921; IGASERPTASE.
CC KW Hydrolase; Serine protease; Transmembrane; Zymogen; Repeat; Signal.
CC FT SIGNAL 1 25
CC FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.
CC FT PROPEP 1015 1702 HELPER PEPTIDE (POTENTIAL).
CC FT ACT_SITE 288 288 PROBABLE.
CC FT DOMAIN 1109 1124 2 X 8 AA TANDEM REPEATS OF A-K-V-E-K-E-E-
CC FT REPEAT 1109 1116 K.
CC FT REPEAT 1117 1124 1.
CC FT REPEAT 1117 1124 2.
CC SQ SEQUENCE 1702 AA; 186539 MW; 860F70D267807A6 CRC64;

alignment_scores:
Quality: 1517.00 Length: 1881
Ratio: 1.651 Gaps: 53
Percent Similarity: 48.857 Percent Identity: 24.615

alignment_block:
US-09-303-518D-651 x IGA2_HAEIN ..
Align seg 1/1 to: IGA2_HAEIN from: 1 to: 1702

64 GCGTTCGCGCGCTGCTTACGATATGCGTCGCGTTCGATTCCTTC 113
1 LysPheLysLeuAsnPhelIEAlaLeuThrValAlaTyAlaLeuThr 21
114 CCAAGCTTGCGGCGGACACACTTATTCGCGATCACTACCAATATCATC 163
1 LysPheLysLeuAsnPhelIEAlaLeuThrValAlaTyAlaLeuThr 21
21 OTYThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38
164 GCGACTTTCGCGGAATTAAGGCAAGTTCAGTCGCGGCGGAAGATAT 213
1 LysPheLysLeuAsnPhelIEAlaLeuThrValAlaTyAlaLeuThr 21
38 rGAspRheAlaGluAsnLysGlyArgPheSerValGlyAlaThrAsnVal 54
214 GAGGTNTACAAACAAAAGGAGGAGTTCGCGCAATCATGACAAAGC 263
1 LysPheLysLeuAsnPhelIEAlaLeuThrValAlaTyAlaLeuThr 21
55 GluValArgAspLysAsnAsnHisSerLeuGlyAsnValLeuProAsnG 71
264 C...CCGATGATTGATTTTCTGTGTGTGCG...CGTACGGCGGTGCGG 307
1 LysPheLysLeuAsnPhelIEAlaLeuThrValAlaTyAlaLeuThr 21
71 YLEProMetIleAspPheSerValValAspValAspTyrArgIleAla 88
308 CATTGGTGGCGGATCAATATATGTGTGAGCGTGCACAT...AACGGCGC 354
1 LysPheLysLeuAsnPhelIEAlaLeuThrValAlaTyAlaLeuThr 21
88 hLEuILeAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104
355 TATAACAACGTGATTTTGTG...GCGGAAGGAAGMA 389
1 LysPheLysLeuAsnPhelIEAlaLeuThrValAlaTyAlaLeuThr 21
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGly 121
390 TCCGATCAGACACCGT...TTTCTTACCAATATG 421
1 LysPheLysLeuAsnPhelIEAlaLeuThrValAlaTyAlaLeuThr 21
121 nAspLysSerHisArgAspValSerSerGluGluAsnArgTyrPheSer 138
422 TGAAGAAGAAATTAATTA...AAG 441
1 LysPheLysLeuAsnPhelIEAlaLeuThrValAlaTyAlaLeuThr 21

```

138 aLgLUyAsnGluTyrProThrLysLeuAsnGlyLysAlaValThrThr 154  
442 CCTGCAATTCACACCTTACACGGCGCATATATGCGCGTTGCA 491  
155 GLuAspGlnThrGlnLysArgArgLysPtyrMetProAlaGLeuAs 171  
492 TAAATTTGTCAAGATGACAGAACTGTGCA...ATGACAGTACATGA 538  
171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaSer 188  
539 GGGGAAATACCTATTCGATTAAGAAAATATCCGAGCGTCCGATC 588  
188 sPAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204  
589 GGCCTGACAGCACCACTATTGGCTTATGATGATGACAAACGCGCATTT 638  
205 GlySerGlySerGlnPheIleTyr.....LysLysGlyAsp.. 216  
639 ATCTACTCCGCGCATGTAAATTTGGCGCAATACATATGACAGGTT 688  
217 .AsnTyrSer.....LeuIleLeuAsnAsnHisGluValGly. 228  
689 GGGCAATTAAT.....GGCGTANTT 708  
229 ..GlyAsnAsnLeuLysLeuValGlyAspAlaTyrThrTyrGlyIleAla 244  
709 AGTTGACGCGCATGTGCGCATCCACAGAC.....TATGG 746  
245 GlyThrProTyrLysValAsnHisGluAsnAsnLysLeuIleGlyPheG 261  
747 C.....C 748  
261 yAsnSerLysGluGlnHisSerAspProLysGlyIleLeuSerGlnAsp 278  
749 CTATGCCGATTCAGAGTCCGCGACGCGACGCGTTCGCCATGTTATT 798  
278 rOleuThrAsnTyrAlaValLeuGlnLysSerGlySerProLeuPheVal 294  
799 TATGCAAAACAAACAAATAAATGGCTGCTCAAGAGTATTACAAACGG 848  
295 TyrAspAlaGlyLysGlyLysTyrLeuPheLeuGly.....Se 307  
849 CTACCTTATTCGCGAGGAGAAACGGTTTCACGCTGATACCAAGATT 898  
307 rTyAspPheThrPAlaGlyTyrAsn.....LysLysSerT 319  
899 GGTTCACGATGACATTTACAGAGCGCATACACATACCGCTNTTTGAA 948  
319 rPGLnGluTyrAsnIleTyrLysProGlu..... 328  
949 CCGCGCATTAACGCGCATTTTCCTTTACATCCAAACAAACGGTAGC.. 996  
329 .....PheAlaLysThrValLeuAspLysAspThrAl 339  
997 .GATACGGTAAACAGAAACCAAGAAAGGTTC...AATCCAAAGCTTA 1042  
339 aGlySerLeuThrGlySerAsnThrGlnTyrAsnTyrAsnProThrGlyL 356  
1043 AAGTACACAGAGTCCGACTGTTTGACGATCTTGAT..... 1080  
356 ysthrSerValIleSerAsnGlySerGlnSerLeuAsnValAspLeuPhe 372  
1081 .....GAACGTATAAGAAACCATTTACGCGGCAAGGGGTGT 1118  
373 AspSerSerGlnAspThrAspSerLys..... 381  
1119 TAATCAGTACGCTCCAGGTTAAACAAACGGTGAACCTTTCTTTATCG 1168  
382 .....LysAsnAsnHisGlyLysSerValThrLeu... 391  
1169 ATTACGCAAGCGCAAACTCATCTTATCAAAACATCAACCAAGCGCG 1218  
392 ..ArgGlySerGlyThrLeuThrLeuAsnAsnIleAspGlnGlyAla 407

1219 GCGCGTTTGTATTGTAAGTGTATTTACGGTC...TCGCGTGAACAA 1265  
408 GLyGlyLeuPhePheGluGlnLysAspTyrGluValLysGlyThrSerAsp 424  
1266 CGAAGCGTGGCAAGCGCGCGCTCATATCACTGAAGACATACCGTTA 1315  
424 rThrThrTyrPLeuGlyAlaGlyAlaSerValAlaAspGlyLysThrVal 441  
1316 CTGGAAAGTAAACGGCGTGGCAACGACCGCTGTCCAAATCGGCAAA 1365  
441 hTrpLysValHisAsnProLysSerAspArgLeuAlaLysIleGlyLys 457  
1366 GGCACGCTGACGTTCAAGCAAGGAGAAACACGCGATTCGATCGCT 1415  
458 GlyThrLeuIleValGluGlyLysGlyLysAsnLysGlySerLeuLysVal 474  
1416 GGGCGACGGTACAGTCATTTGGATCAGCAGCAGCAGATAAAGCAAA 1465  
474 lGlyAspGlyThrValIleLeuLysGlnGlnAlaAspAlaAsnLysVal 491  
1466 AACAGCCTTTAGTGAATCGCTTGNTACGCGCAGGCGTACGTTGCA 1515  
491 aLysAlaPheSerGlnValGlyIleValSerGlyArgSerThrValVal 507  
1516 CTGAATGCCGATTAATCACTTCAACCGCGCATTCGCTTCGTCACCGTATC 1565  
508 LeuAsnAspAspLysGlnValAspProAsnSerIleTyrPheLysPheArg 524  
1566 GCGCGGACGTTTGAATTAACCGCGCATTCGCTTCGTCACCGTATC 1615  
524 gGlyGlyArgLeuAspAlaAsnGlyAsnAsnLeuThrPheGlnHisIleA 541  
1616 AAAATACCGATGAAGCGCGATATTGCNATATATATGCGACACAAACA 1665  
541 rGAsnIleAspAspGlyAlaArgLeuValAsnHisAsnThrSerLysThr 557  
1666 TCCACGCTTACATTAACGAGGAAAGTAAAGTATTAACACCGAGT... 1710  
558 SerThrValThrIleThrGlyGlnSerLeuIleThrAspProAsnThrI 574  
1710 ..... 1710  
574 eThrProTyrAsnIleAspAlaProAspGluAspAsnProTyrAlaPheA 591  
1710 ..... 1710  
591 rGArgIleLysAspGlyGlyGlnLeuTyrLeuAsnLeuGlnAsnTyrThr 607  
1710 ..... 1710  
608 TyrTyrAlaLeuArgLysGlyAlaSerThrArgSerGluLeuProLysAs 624  
1711 .....GGTAGAAT..... 1719  
624 nSerGlyGlnSerAsnGluAsnThrPLeuTyrMetGlyLysThrSerAspG 641  
1720 .....ATCAATAGACTTAATTACGCAAGAAATATGCGC 1752  
641 lualLysArgAsnValIleMetAsnHisIleAsnAsnGlnIleArgMetAsnGly 657  
1753 TACACGGTGGTGTGGCGAAGAAATACGACCAAAACGAGCGGGGCT 1802  
658 PheAsnGlyTyrPheGlyGluGlnGlu...GlyLysAsnAsnLysAsnIe 673  
1803 CAACCTTGTTTACACCGCGCGCAGAGACGCGACCGCTGCTTCCG 1852  
673 uAsnValThrPheLysGlyLysSerGlnGlnAsnArgPheLeuThrG 690  
1853 GCGGAACAATTTAAACGGCAACATCACGCAAAACAAACAACTGTT 1902  
690 LysGlyThrAsnLeuAsnGlyAspLeuLysValGlnLysGlyThrLeuPhe 706

```
1903 TTCAGCGGAGACCGACCGACCGCTTACATCAATTGAGAGCGGGTG 1952
      ::::::::::::::::::::
707 LeuSerGlyArgProThrProHisAlaArgAspIleAlaGlyIleSer 723
      ::::::::::::::::::::
1953 GTCAAAAATGGAAGT.....ATCCACAGAGAGAAATCGTGGGACA 1996
      ::::::::::::::::::::
723 rThrLysLysAspGlnHisPheAlaGlnAsnGluValValGluA 740
      ::::::::::::::::::::
1997 ACAGCGATGTCNACGCGAGCTTAAAGCGGAAATTCATATTCAGGCG 2046
      ::::::::::::::::::::
740 spAspTrpIleAsnArgAsnPheLysAlaThrAsnIleAsnValThr 756
      ::::::::::::::::::::
2047 GGGCGAGCGGTGATTTC...CGCAATGTGTCCAAAGTGAAGCGGATG 2093
      ::::::::::::::::::::
757 AsnAlaThrLeuTyrSerGlyArgAsnValAlaAsnIleThrSerAsn 773
      ::::::::::::::::::::
2094 NCATTGGACATCAGCGCCAGCAGTTTGTGTGCGACCGCATCAA 2143
      ::::::::::::::::::::
773 eThrAlaSerAspAsnAlaLysValHisIleGlyTyr.....LysAlaG 788
      ::::::::::::::::::::
2144 GCCATCAATCTGTACAGCTGCGACTGGACNGCTGTGACAAATTGTGC 2193
      ::::::::::::::::::::
788 LysPheValCysValArgSerAspTyrThrGlyTyrValThrCysThr 804
      ::::::::::::::::::::
2194 GAANAANCAATTACCGACGATAAAGTATGCTTCATTGACTAAGACNGA 2243
      ::::::::::::::::::::
805 ThrAspLysLeuSer...AspLysAlaLeuAsnSerPheAsnAlaThrAs 820
      ::::::::::::::::::::
2244 CATTAGCGGCAATGTAGNCTNCCNATAGCNTNTTAAANCTCNGC 2293
      ::::::::::::::::::::
820 nValSer..... 822
      ::::::::::::::::::::
2294 GGCNTCCNNCACTNNAAGCAATCTTAGCAATGGCGATACGTTAT 2343
      ::::::::::::::::::::
822 ..... 822
      ::::::::::::::::::::
2344 ACAGTCAGCCACACGCCACCCAAAACGGCAACTTAGCCTGTGGGCA 2393
      ::::::::::::::::::::
823 .....GlyAsnValAsnLeuSerGlyAs 830
      ::::::::::::::::::::
2394 TGCCCAAGCACTTATTAATCAAGCCACATTAAGCGCAACNCATGGNT 2443
      ::::::::::::::::::::
830 nAlaAsnPheValLeuGlyLysAlaAsnLeuPheGlyThrIleSerGlyT 847
      ::::::::::::::::::::
2444 CGGCGCAATGCTCATTTATCTAAGCAACCGCCGCAACAAAAGCGAGT 2493
      ::::::::::::::::::::
847 hGlyAsnSerGlnValArgLeuThrGluAsn..... 857
      ::::::::::::::::::::
2494 CTGACGCTTCCGACAAAGCTAAGCAAAAGTAAAGCATTCGCGACGAA 2543
      ::::::::::::::::::::
857 ..... 857
      ::::::::::::::::::::
2544 CGGCAATGCTCCTTAGCCGATAGGCAAGTATTCATTTGAAAACAGCC 2593
      ::::::::::::::::::::
857 ..... 857
      ::::::::::::::::::::
2594 GCTTACCGGACACTCAGCGGAGAGAGANACAGATTACACTTAAA 2643
      ::::::::::::::::::::
857 ..... 857
      ::::::::::::::::::::
2644 GACAGCAATGAGCGCTGCGCTCAGGACGAGAAATTAGCAATTAAACCT 2693
      ::::::::::::::::::::
858 ...SerHisTrpHisLeuThrGlyAspSerAsnValAsnGlnLeuAsnIle 873
      ::::::::::::::::::::
2694 TGAACAAGCCACATTTACATCAATTCGCGCTATGCGACGATGCTGAG 2743
      ::::::::::::::::::::
873 uAspLysGlyHisIleHisLeuAsnAla.....GlnAsnAspAlaAsnL 888
      ::::::::::::::::::::
2744 GGGCGAAACCGGCAACNGTGTACAGACCGCGCGCGCGCTGCGCGCT 2793
      ::::::::::::::::::::
888 yValAlaThr..... 891
      ::::::::::::::::::::
2794 TCCCTATTATCCGTTACACCGGCAACTTCGTAAGTAACCGGTTCAAC 2843
      ::::::::::::::::::::
      ::::::::::::::::::::
892 .....TyrAsnThr 894
      ::::::::::::::::::::
2844 GCTACGCGTAAACCGCAATTTGAACNGTCAAGAGACATTCGCGTTATCT 2893
      ::::::::::::::::::::
894 rLeuThrValAsn...SerLeuSerGlyAsnGlySerPheTyrTyrLeuT 910
      ::::::::::::::::::::
2894 CGGAACCTCTTCGCGTACCGAAGCGACAAATTTGAAGCTGGCGGAAAGTTC 2943
      ::::::::::::::::::::
910 hAspLeuSerAsnLysGlnGlyAspLysValValAlaThrLysSerAla 926
      ::::::::::::::::::::
2944 GAAGNACTTACACCTTGGCGGTCAACATACCGGACAGACACCGCTAAG 2993
      ::::::::::::::::::::
927 ThrLysAsnPheThrLeuGlnValAlaAspLysThrGlyLupProThrL 943
      ::::::::::::::::::::
2994 CCTGATCATATTGACGCTAGTGGAGGAGGAAAGCAACAAACCCGTGCG 3043
      ::::::::::::::::::::
943 s...AsnGluLeuThrLeuPheAspAla.....SerAsnAlaThrArg 957
      ::::::::::::::::::::
3044 AAACCTTAATTTCCACCTGCAAAACGACACGTCATGCGCGCGGTG 3093
      ::::::::::::::::::::
957 snAsnLeuAsnValSerLeuValGlyAsnThrValAlaPheLysAlaTrp 973
      ::::::::::::::::::::
3094 CGTTACCACTCATCCGCAAAAGCGGAGTTCGCTGCATATTCG.. 3141
      ::::::::::::::::::::
974 LysTyrLysLeuArgAsnValAsnGlyArgTyrAspLeuTyrAsnProG 990
      ::::::::::::::::::::
3141 ..... 3141
      ::::::::::::::::::::
990 uValGluLysArgAsnGlnThrValAspThrThrAsnIleThrThrPro 1007
      ::::::::::::::::::::
3141 ..... 3141
      ::::::::::::::::::::
1007 snAsnIleGlnAlaAspValProSerValProSerAsnAsnGluIle 1023
      ::::::::::::::::::::
3141 ..... 3141
      ::::::::::::::::::::
1024 AlaArgValGluThrProValProProProAlaProAlaThrProSer 1040
      ::::::::::::::::::::
3141 ..... 3141
      ::::::::::::::::::::
1040 uThrThrGluThrValAlaGluAsnSerLysGlnGluSerLysThrVal 1057
      ::::::::::::::::::::
3141 ..... 3141
      ::::::::::::::::::::
1057 LuLysAsnGluGlnAspAlaThrGluThrThrAlaGlnAsnGlyGluVal 1073
      ::::::::::::::::::::
3141 ..... 3141
      ::::::::::::::::::::
1074 AlaGluGluAlaLysProSerValLysAlaAsnThrGlnThrAsnGlu 1090
      ::::::::::::::::::::
3142 .....GTCAAG 3148
      ::::::::::::::::::::
1090 LAlaGlnSerGlySerGluThrGlnGluThrGlnThrGluIleLysG 1107
      ::::::::::::::::::::
3149 AACAGAGCTTTCGACAAACTCGGACAGCGCAGAGCAAAAGAAACAG... 3195
      ::::::::::::::::::::
1107 LuThrAlaLysValGluLysGluLysLysAlaLysValGluLysGluGlu 1123
      ::::::::::::::::::::
3195 ..... 3195
      ::::::::::::::::::::
1124 LysAlaLysValGluLysAspGluIleGlnGluAlaProGlnMetalase 1140
      ::::::::::::::::::::
3195 ..... 3195
      ::::::::::::::::::::
1140 rGluThrSerProLysGlnAlaLysProAlaProLysGluValSerThra 1157
      ::::::::::::::::::::
3195 ..... 3195
      ::::::::::::::::::::
1157 spThrLysValGluGluThrGlnValGlnAlaGlnProGlnThrGlnSer 1173
      ::::::::::::::::::::
3195 ..... 3195
      ::::::::::::::::::::
```

```

1174 ThrThrValAlaAlaIaIaIaThrSerProAsnSerLysProAlaGI 1190
3196 .....GGGAAAAAGAACGGCAAGCCTTGACCGGCTGA 3232
1190 uGIuThrcInProSerGIuLysThrAsnAlaGIuProValThrProValY 1207
3233 TTGCGCGCGGG..... 3243
1207 aLserLysAsnGIuThrcIuAsnThrThrAspGIuProThrGIuArgIu 1223
3244 CGCGATGCCCGCGCAAGACAGAAAGCCTTGCCAGACCGCGCGCGCGC 3293
1224 LysThrAlaLysValGIuThrGIuLysThrGIuInuProGIuValAl 1240
3294 AGGC..... 3297
1240 aSerGIuAlaSerProLysGIuInuGIuInSerGIuThrValGIuProGIuA 1257
3298 .....GGGAAAAATGTC...GGCATTATGACAGCGGAGAG 3333
1257 LaValleuGIuSerGIuAsnValProThrValaAsnAlaGIuGIuVal 1273
3334 AAAAAGCGGTGCAGCGCGATAAAGC..... 3360
1274 GIuAlaGIuInuGIuInThrGIuSerAlaThrValSerThrLysGIuPr 1290
3361 .....AGCGCNTTGGCGAAGC 3376
1290 cAlaProGIuAsnSerIleAsnThrGIuSerAlaThrAlaIleThrGIuT 1307
3377 AGCGCGAAGCGAAACCGCGCGGNTAC..... 3405
1307 hrAlaGIuLysSerAspLysProGIuInThrGIuThrAlaIaIaSerThrGIu 1323
3405 ..... 3405
1324 AspAlaSerGIuInHsLysAlaAsnThrValAlaAspAsnSerValAlaAs 1340
3406 .....ACCGCTTCCCGCGCGCGCGCGCGCGGATGTC 3445
1340 nAsnSerGIuSerGIuProLysSerArg...ArgArgArgSerLies 1356
3446 CGCAACCGCAC..... 3456
1356 erGIuProGIuInuThrSerAlaGIuInuThrAlaIaIaSerThrAsp 1372
3457 .....CCCAACCGCAACCTCAACCCCAAC 3482
1373 GIuThrThrIleAlaAspAsnSerLysArgSerLysProAsnArgArgSe 1389
3483 GCAG..... 3486
1389 rArgArgSerValArgSerGIuProThrValThrAsnGIuSerAspArgS 1406
3487 .....CGGACCTGATNAGCGGTATGCCAATAGCGGTTTG 3522
1406 erThrValAlaIaIaLeuArgAspLeuThrSerThrAsnThrAsnAlaValIle 1422
3523 AGTGAATTTCCGCC.....ACGCTCAACAGCGTTTCCG 3557
1423 SerAspAlaMetAlaIaLysAlaGIuInPheValAlaIaLeuAsnValGIuLysAl 1439
3558 CGTACAGACGAATTTGACCGCGT...TTTGGCGAAGACCGCGCGCAACG 3604
1439 aValSerGIuInHsIleSerGIuInuGIuMetAsnAsnGIuGIuInuYrA 1456
3605 CNGTTTGACAGCNGCATCCGGNACCAAAACACTACCGTTGCCAAGAT 3654
1456 snValITrPValSerAsnThrSerMetAsnGIuAsnTYrSerSerSerGIu 1472
3655 TTTCGGCGCTTACCGCAACAAACGACCTGGCCCAATCGGTATGCGAA 3704
1473 TyrArgArgPheSerSerLysSerThrGIuInThrGIuInuGIuTYrPaspGI 1489

```

```

3705 AAMCTCGGACGGG...CGCGTGGCATTCCTGTTTGGCAACCCGA 3751
1489 nThrIleSerAsnAsnValGIuInuGIuGIuValPheThrValArgA 1506
3752 CCGAAACANCTTTCGACGAGCGATCGCACTCGCAGCGCTGCCAC 3801
1506 snSerAsnAsnPheAspLysAlaSerSerLysAsnThrIleuAlaInVal 1522
3802 GCGCGCGTTTGGGCATACGGCATCGCAGGTTTCAGATCGGCATCAG 3851
1523 AsnPheTYrSerLysTYrTYrAlaAspAsnHsITrTYrLeuGIuL 1539
3852 CACGGCGCGGTTTTCAGCGCGCANCTTCACAGCGCATCGCAGCA 3901
1539 PleuGIuTYrGIuLysPheGIuInSerAsnLeuLysThrAsnHsAlaI 1556
3902 AATTCGCGCGCGCTGCTGCATTACGGCATTCAGCAGCATACCGCGC 3951
1556 YspHeAlaIaIaGIuHsIThrAlaGIuInPheGIuLeuThrAlaGIuLysAlaPhe 1572
3952 GGTTCGGCGGATTCGCGCATCGACCGTACATCGCGCAACCGCTATT 4001
1573 AsnLeuGIuLysnPheGIuIleThrProIleValGIuValArgTYrSerTY 1589
4002 CGTCCAAAAAGCGGATTACCGGTACGAAACGTCATATCGCACCCCG 4051
1589 rLeuSerAsnAlaAsnPheAlaIaLeuAlaLysAspArgIleLysValAsn 1606
4052 GTCTGCGTTCAACCGTACCGCGCGGATTAAGCGATATTCATTC 4101
1606 rIleSerValLysThrAlaPheAlaGIuValAspLeuSerTYrThrTYr 1622
4102 AAMCGCGCAACACATNTCCATACNCCTTATTAACTGTCTCAT... 4149
1623 HsLeuGIuGIu...PheSerValThrProIleLeuSerAlaArgTYrAs 1638
4150 .ACCGATCGCGCTTCGGCAAGTCCGACACGCGTCATATTCGCGGTAT 4198
1638 pThrAsnGIuInuLysSerGIuLysIle....AsnValAsnGIuTYrAsp 1653
4199 TGCTCTAGATTTGCGCAAAACCGCGAGTGGCAATGGGCGCTAAACGCC 4248
1653 heAlaTYrAsnValGIuAsnGIuInuGIuTYrAsnAlaGIuLysLeu 1669
4249 GAAATCAAGGTTTCACGCTGTCCNTCCACGCTGCCCGCGCCCAAGNCC 4298
1670 LysTYrHsAsnValLysLeuSerLeuIleGIuGIuLysLeuThrLysAla 1686
4299 GCAACTGGAAGCGCACACAGCGCGGCAATCAATATGAGCTAC 4341
1686 sgInAlaGIuLysGIuLysThrAlaGIuLeuLysLeuSerPhe 1700

```

seq\_name: SwissProt\_40:IGA\_NEIGO

```

seq_documentation_block:
ID IGA_NEIGO STANDARD; PRT; 1532 AA.
AC P09790:
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Iga-specific serine endopeptidase precursor (EC 3.4.21.72) (Iga
DE protease).
GN IGA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBL_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=MS11;
RX MEDLINE=87115823; PubMed=3027577;
RA Pohlner J., Hatter R., Beyreuther K., Meyer T.F.;
"Gene structure and extracellular secretion of Neisseria gonorrhoeae

```

RT Iga protease.";  
 RL Nature 325:458-462(1987).  
 RN [2]  
 RP ACTIVE SITE:  
 MEDLINE=90154052; PubMed=2105953;  
 RA Bachovich W.W., Plant A.G., Fleitke G.R., Lynch M., Kettner C.A.;  
 RT "Inhibition of Igal proteinases from *Neisseria gonorrhoeae* and  
 RT Hemophilus influenzae by peptide prolyl boronic acids";  
 RL J. Biol. Chem. 265:3738-3743(1990).  
 CC -1- FUNCTION: THIS PROTEASE IS SPECIFIC FOR IMMUNOGLOBULIN A.  
 CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at  
 CC certain Pro-I-Xaa bonds in the hinge region. No small molecule  
 CC substrates are known.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDE THE PRECURSOR TO THE PERIPLASMIC  
 CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE  
 CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE  
 CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X04835; CAA28538.1; -  
 DR PIR: A26039; A26039.  
 DR MEROPS: S06.001; -  
 DR InterPro: IPR000710; IGA\_S6.  
 DR Pfam: PF02395; IGA1; 1.  
 DR PRINTS: PRO0921; IGASERPTASE.  
 KW Hydrolase; Serine protease; Zymogen; Autocatalytic cleavage;  
 KW Transmembrane; Signal.  
 FT SIGNAL 1 27  
 FT CHAIN 28 986 IGA-SPECIFIC SERINE ENDOPEPTIDASE.  
 FT PROPEP 987 1532 HELPER PEPTIDE.  
 FT ACT\_SITE 278 278 POTENTIAL.  
 FT SITE 986 987 CLEAVAGE (AUTO-).  
 FT SITE 1018 1019 CLEAVAGE (AUTO-).  
 FT SITE 1121 1122 CLEAVAGE (AUTO-).  
 SQ SEQUENCE 1532 AA; 168976 MW; 68FF4112BD22F40D CRC64;

## alignment\_scores:

Quality: 1508.00 Length: 1727  
 Ratio: 1.735 Gaps: 45  
 Percent Similarity: 50.318 Percent Identity: 26.288

## alignment\_block:

US-09-303-518D-651 x IGA\_NEIGO ..

Align seg 1/1 to: IGA\_NEIGO from: 1 to: 1532

49 AAACGGCGGCGATCGCTTCGCGCTTACTTACCATATGCGCTTC 98  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 2 LysAlaLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18  
 99 GTTCGGCATTTCTCCCAAGCTTGGCGGGGACACACTATTTCGGCATCA 148  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 18 alyrAlaLeuThrProTyrSerGluAlaIleValaIleValaIleVala 35  
 149 ACTACCAATTCATGCGACTTTCGCGAAATAAAGCAAGTTCGCACTC 198  
 ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
 35 sPTyGlnIlePheArgPheAlaGluAsnLysGlyLysPhePheVal 51  
 199 GGGGCGAAGATATTGAGGTNTACAACAAAAGGGAAGTTCGCGCAA 248  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 52 GAlaIleThrAspLeuSerValLysAsnLysArgGlyGlnAsnIleGlyAs 68  
 249 ATCAATGACAAAAGCCCGATGATTGTTTCTGTGTGTCGCGTAAC. 297

68 nAlaLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnL 85  
 ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
 298 ..GGCGTGGCGGCGATCGTGGCGATCAATATTCGTCGAGCGGCAAT 345  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 85 ysArGllleAlaThrValValaIleAspProGlnAlaValSerValLysHis 101  
 346 .....AACGGCGGTATACACGCTTGA 368  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 102 AlAlaLysAlaGluValaHisThrPheTyrTyrGlyGlnTyrAsnLysHis 118  
 369 TTTTGTGGCGAAGGAAGNATCCCATAGCACCGCTTTTCTTACCAA 418  
 ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
 118 nAspValAlaAspLysGluAsn.....GluTyrArgY 129  
 419 TTGGAAGAAGAAATATTAATACCTGCATTCACACCCCTTACACGGG 468  
 ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
 129 alValGluGlnAsnAsnTyrGluProHisLysAlaTrpGlyAlaSerAsn 145  
 469 .....GATTCACATATGCCGCTTTCATTAATTTGTCAC 503  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 146 LeuGlyArgLeuGluAspTyrAsnMetAlaArgPheAsnLysPheValh 162  
 504 AGATGCAAACTGTGGAATGACAGTACATGAGGGG...AATACCT 550  
 ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
 162 rGluValAlaProIleAlaProThrAspAlaGlyGlyLeuAspThrT 179  
 551 ATTCCGATTAAGAAAATATATCCGAGCGTCCGCAATGCCGTGAGACAC 600  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 179 yryAspLysAsnArgPheSerSerPheValArgIleGlyAlaGlyArg 195  
 601 CAC.....TATTGGCTTATGATGATGACAAACCGG 632  
 ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
 196 GlnLeuValTyrGluLysGlyValTyrHisGlnGlyAsnGluLysG 212  
 212 yTyrAspLeuArgAspLeuSerGlnAlaTyrArgTyrAlaIleAlaGlyT 229  
 670 .....AATACAT...ANG 681  
 229 hrProTyrLysAspIleAsnIleAspGlnThrMetLsnThrGlyLeu 245  
 682 CAGGGTGGGGAATAATAGCGGTANTTACTTGGCGGCGATGGCGCA 731  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 246 IleGlyPheGlyAsnHisAsnLysGlnTyrSerAlaGluLeuLysG 262  
 732 TGCCACGACTATGGCCCTATGCCGATGAGTGGCGGCGGACAGCG 781  
 ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
 262 nAlaLeuSerGlnAspAlaLeuThrAsnTyrGlyValLeuGlyAspSerG 279  
 782 GTTCGCAATGTTTATTTATGACAAACAAACAATAATAGTGTGTCAC 831  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 279 LysErProLeuPheAlaPheAspLysGlnLysAsnGlnTrpValPheLeu 295  
 832 GGAAGTTTACAACCGGCTACCTTATTCGCGGAGGAAACGGTTTCCA 881  
 ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
 296 GllTyrThrAspTyrTrpAlaGlyTyrGly..... 305  
 882 GCTGATACGCAAGATTTGTTTACGATGACATTTACAGAGCGAT... 927  
 ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||  
 306 .....LysLysSerTrpGlnIleTrpAsnIleTyrLysLysGlnPheA 320  
 928 .....ACACATACCGTNTTTTGAACCGCGCATACGCA 963  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 320 IAspLysIleLysGlnHisAspAsnAlaGlyThrValLysGlyAsnGly 336  
 964 .....CATTTTCCTTACATCAACAACAACGAGTACGGTAAAC 1007  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 337 GlnHisIleTrpLysThrThrGlyThrAsnSerHisIleGlySerThr... 352  
 1008 AGAACCACAGCAAAAGTNTCCATCCAAAGCTTAAGTACAGACAGTGC 1057  
 ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

```
353 .....Alavala 355
1058 GACTGTTGACGAATCTTTGAATGAAGTAAAGAACAGTTTACGCG 1107
      |||||
355 rgleuAlaasn.....AsnGluGlyAsp..... 362
1108 GCAGGGGGGTGTAATCAGTACCGTCCAAAGGTTAAACAAAGGTGAACCT 1157
      |||||
363 .....AlaasnAsnGlyGlnAsnVa 369
1158 TTCTTTATCGATTACGGCAAGCGCAACTCATCTTATCAACAAACATCA 1207
      |||||
369 lthrPheGluAsp.....AsnGlyThrLeuValLeuAsnGlnAsnIleA 384
1208 ACCAAGCGCGCGGCGTTGTATTTGAAGTATTTACGTTTCACGCTCCGCT 1257
      |||||
384 snGlnGlyAlaGlyLeuPhePhePheGlyAspTyrThrValLysGly 400
1258 GAAACAAACGAA...ACGTGGCAAGCGCGCGGCTTCATATCAGTGAAGA 1304
      |||||
401 AlaasnAsnAspIleThrThrLeuGlyAlaGlyIleAspValAlaAspG1 417
1305 CAGTACCGCTTACTGTAAGAAAGTAAACGGCGTGCAACGACGCGCTGTCCA 1354
      |||||
417 TyrLysValValAlaThrGlnValLysAsnProAsnGlyAspAlaAla 434
1355 AATCGGCAAGGCAAGCGCTGACGTTCAAGCCAAAGGGGAAACCAAGGC 1404
      |||||
434 ysIlleGlyGlyThrLeuGlnIleAsnGlyThrGlyValAsnGlnGly 450
1405 TCGATCAGCGTGGCGGCGGTACGCTACGCTGATTTGATCAGCAGCGACGA 1454
      |||||
451 GlnLeuValGlyAlaAspLysPheLysValIleLeuAsnGlnLysAlaAspAl 467
1455 TAAAGCAAAACAAAGCCTTACTGTAATGCGCTTGTGNTCAGCGGCAAGG 1504
      |||||
467 AspLysLysValGlnAlaPheSerGlnValGlyLeuValSerGlyArg 484
1505 GTACGGTGAAGTGAATGCGGTAATACGTTCAACCGGACAACTGAT 1554
      |||||
484 lYThrLeuValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeuTyr 500
1555 TTCGGCTTTCGGCGGCGGAGTTTGATTTAAACGGGCAATTCGCTTCGCT 1604
      |||||
501 PheGlyPheAlaGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThrPh 517
1605 CCACCGTATTCAAATACGATGAGGCGGCGATGATTCNCATCATATG 1654
      |||||
517 eGlnHisIleArgAsnValAspGluGlyAlaArgIleValAsnHisAsnT 534
1655 CCACAACAACATCCACCGTACCATTCAGAGGATGAAGATATTATTCACAA 1704
      |||||
534 hAspHisAlaSerThrIleThrLeuThrGlyLysSerLeuIleThrAsn 550
1705 CCGAGT..... 1710
551 ProAsnSerLeuSerValHisSerIleGlnAsnAspTyrAspGluAspAs 567
1711 .....GTTAAGATA 1720
567 pTySerTyrTyrTyrArgProArgArgProIleProGlnGlyLysAspL 584
1721 TC.....AATGACTT 1731
      |||||
584 euTyrTyrLysAsnTyrArgTyrTyrAlaLeuLysSerGlyArgLeu 600
1732 AAT..... 1734
      |||||
601 AsnAlaProMetProGluAsnGlyValAlaGluAsnAsnAspTyrPhePh 617
1735 .....TACAGCAAGAATTTGCC..... 1752
      |||||
617 emetGlyTyrThrGlnGluAlaArgLysAsnAlaMetAsnHisLysA 634
1753 .....TACAAAGTGTGTGGCGAAGAAATAGC 1782
      |||||
634 snAsnArgArgIleGlyAspPheGlyLysPhePheAspGluGluAsnGly 650
1783 ACCAAACGAACGGCGGCTCAACCTTGTATACAGCGCGCGCAGAGA 1832
      |||||
651 LysGlnHisAsnGlyAlaLeuAsnLeuAsnPheAsnGlyLysSerAlaG1 667
1833 CCGCACCCGCGTGTCCCGCGAACAATTTAAAGCAACATCAGC 1882
      |||||
667 nLysArgPheLeuLeuThrGlyAlaAlaAsnLeuAsnGlyLysIleSerV 684
1883 AAACAAAGCGCAACTGTTTTCAGCGGACGACGACGACGCGCTTAC 1932
      |||||
684 alThrGlnGlyAsnValLeuLeuSerLysArgProThrProHisAlaArg 700
1933 AATCATTTAGAACGGGTGCTCAAAAATGGAAGT.....ATCCACA 1976
      |||||
701 AspPheValAsnLysSerSerAlaArgLysAspAlaHisPheSerLysAs 717
1977 AGGAGAAATGCTGCGGACAAACGCTGATCMCCGACGTTTAAAGCG 2026
      |||||
717 nasGlnValValPheGluAspAspTyrPheAsnArgThrPheLysAla 734
2027 AAATTCATAT.....CAGCGCGGCGAGCGGTGATTTCCCGCATGT 2073
      |||||
734 lalGluIleAlaValAsnGlnSerAlaSerPheSerSerGlyArgAsnVal 750
2074 GCCAAAGTGGAGCGCATTTGNCATTTAGCAATACAGCCCAAGCAGTTT 2123
      |||||
751 SerAspIleThrAlaAsnIleThrAlaThrAspAsnAlaLysValAsnLe 767
2124 TGGTGTGCGACCGCATCAAGCCATACACTGTGACAGTTGCGAGTGA 2173
      |||||
767 uGlyTyr.....LysAsnGlyAspGluValCysValArgSerAspTyr 782
2174 CNGGTCTGACAAATTTGTGTGAAANAAACATTCACGACGTAAGTAT 2223
      |||||
782 hGlyTyrValThrCys..... 787
2224 GCTTCATTGACTAAGACGACGACNTNAGCGGACANTGNACTNNCNAATNA 2273
      |||||
787 ..... 787
2274 CGNTNTTNAANCTCNCNGCGCTGCGNCACTNAAGGCAATCTTAGTG 2323
      |||||
788 .....AsnThrGlyAsnLeuSer..... 793
2324 CAAATGGCATACAGTTATACAGTCAGCCACACAGCCACCAA...AAC 2370
      |||||
794 .....AspLysAlaLeuAsnSerPheAspAlaThrArgIleAsn 806
2371 GGCACCTTGAAGCTCGTGGGCAATGCCCAAGCAACATTTATATAGCCAC 2420
      |||||
807 GlysAsnValAsnLeuAsnGlnAsnAlaLeuValLeuGlnLysAlaAl 823
2421 ATTAAACGGCAACNCATCGGNTTCGGGCAATGCTTATTAATCTAAGCA 2470
      |||||
823 aleuThrPheGlyLysIleGlnGlyGlnGlyAsnSerArgValSerLeuAsnG 840
2471 ACAAGCGCGCAAAAACGGAGTGTGACGCTTTCGACAAACGCTAAGCA 2520
      |||||
840 lnhIs..... 841
2521 AACGTAAAGCATTCGCGACTCAAGCGCAATGTCTCCCTAGCGATAGCG 2570
      |||||
841 ..... 841
2571 AGTATTCATTTGAAACACAGCGCTTACCGGACAACTCAGCGGACAGA 2620
      |||||
841 ..... 841
```

2621 AGGAAAGACATTTACACTTAAAGACAGCATGCGCCTGCCACGCG 2670  
||||| ||| |||  
842 ..... SerIystrpHisLeuThrGlyAsp 849  
2671 ACGGAATTAGCAATTTAAACCTTGACAGCGCACCATTTACATCATTC 2720  
||||| ||| |||  
850 SerGlnValHisAsnLeuSerLeuAlaAspSerHisIleHisLeuAsnAs 866  
2721 CGCCTATGCCCGCAGATGTCGAGCGCGCAAAACCGCAGNGTGCAGACA 2770  
||||| ||| |||  
866 nAla..... SerAspAlaGlnSerAla..... 873  
2771 CGCGCGCGCGCGCTTCCGCGCGCTTCCATATCCGTTACACCGCAACT 2820  
873 ..... 873  
2821 TCGGTAGAAATCCCGTTTACACGCTGACGGTAAACGGCAATTGACNG 2870  
||||| ||| |||  
874 ..... AsnIystrHisThrIleLysIleAsn... HisLeuSerG1 886  
2871 TCAAGGAACATTCGCTTATGTGCGAACTCTCGGCTACGAGCGACA 2920  
||||| ||| |||  
886 yAsnGlnHisPheHisIyLeuThrAspLeuAlaLysAsnLeuGlyAspL 903  
2921 AATTGACGCTGGCGAAAGTTCCAGAGNACTTACACCTTGCGGCTAC 2970  
||||| ||| |||  
903 yValIleuValLysGlnSerAlaSerGlyHisIyGlnLeuHisValGln 919  
2971 AATACCGGACAGAACCC..... GTAAGCCTCGATCAATTGACGGTGT 3014  
||||| ||| |||  
920 AsnIystrHisGlyLuproAsnGlnIuGlyLeuAspLeuPheAspAlaSe 936  
3015 GGAAGGAGAGACACAAACCCGCTGCCGAAACCTTAATTTCACCCGCG 3064  
||||| ||| |||  
936 rSerValGlnAspArg..... SerArgLeuPheValSerLeuA 949  
3065 AAACGACACGCTGCATGCGCGCGGCTTACCAACTCATCCGCAAA 3114  
||||| ||| |||  
949 lAsnHisIyValAspLeuGlyAlaLeuArgIyThrIleLysThrIu 965  
3115 GACGCGGAGTTCCGCTGCATTAAT..... CCGGT 3143  
||||| ||| |||  
966 AsnGlyIleThrArgLeuYAsnProGlyAlaGlyAsnGlyArgProVa 982  
3144 CAAGACAGACAGACTTCCGCAAACTCGCGCAAGCAGACGCAAAAAAC 3193  
||||| ||| |||  
982 lLysProAlaProSerProAlaAlaAsnThrAlaSerGlnAlaGlnLysA 999  
3194 AGCGGGAAGAACACACGCG..... 3213  
999 lAthrGlnThrAspGlyAlaGlnIleAlaLysProGlnAsnIleValAl 1015  
3214 ..... CAAGCCTTGACGCGCGATTCGGCG 3239  
1016 AlAProToserProGlnAlaAsnGlnAlaGlnIuGlnAlaLeuArgGlnG1 1032  
3240 CGGCGCGATTCGCCGCGCAAAAGACAGAAAGCTGCGCAACCGCGCGCG 3289  
||||| ||| |||  
1032 nAlaLysAlaGlnIuGlnValLysArgGlnIleAlaAlaGlnAlaGlnLysV 1049  
3290 NCGCAGCGCGGGAATGTCGCAATTATGCGACGCGGAGAGAGAAAAA 3339  
||||| ||| |||  
1049 aAla..... ArgGlnLysAspIuGlnAlaLys 1058  
3340 CGGGTCCAGCGGAT..... 3354  
1059 ArgLysAlaAlaGlnIleAlaArgGlnGlnIuGlnAlaArgLysAlaAl 1075  
3355 ..... AAAGACACGCGNT 3367  
1075 aGluLeuAlaAlaLysGlnLysAlaGlnAlaGlnLysAlaArgLysAlaArgGlnL 1092  
3368 TGGCGAAACAGCGCGAAGCGGAAACCGCGCGGNTACACGCGCTTCCCG 3417

||||| ||| |||  
1092 euAlaArgGlnLysAlaGlnIuAlaSerHisGlnAlaAsnAlaLysPro 1108  
3418 CGCGCGCGCGCGCGCGCGCGGATTTGCCGCAACCG..... 3453  
1109 LysArgThrArgAlaAlaIleLeuProArgProProAlaProValPh 1125  
3453 ..... 3453  
1125 eSerLeuAspAspIyThrAlaLysAspAsnSerGlnUserSerIleGlyA 1142  
3453 ..... 3453  
1142 snLeuAlaArgValIleProArgMetGlyArgGlnLeuIleAsnAspIy 1158  
3453 ..... 3453  
1159 GlnGluIleProLeuGlnIuLeuGluAspGluAlaGlnIuGlnArg 1175  
3453 ..... 3453  
1175 gGlnAlaThrGlnPheHisSerLysSerArgAsnArgAlaIleSers 1192  
3453 ..... 3453  
1192 eArgLuproSerAspGluAspAlaSerGlnSerValSerThrSerAsp 1208  
3453 ..... 3453  
1209 LysHisProGlnAspAsnThrGlnLeuHisGlnLysValGlnThrAlaG1 1225  
3454 ..... CAACGCAACCTCAACCCCAACCGCAGCGCG 3490  
1225 yLeuGlnProArgAlaAlaGlnIuProArgThrGlnAlaAlaAlaGlnAla 1242  
3491 ACCTGATNACCGCTTATGCCATAGCGGTTTGATGTAATTTCCGCGACG 3540  
1242 spAlaValSerThrAsnThrAsnSerAlaLeuSerAspAlaMetAlaSer 1258  
3541 CTCACACGCTT..... TTGCGCTACGAGCAAAAT 3572  
1259 ThrGlnIleLeuLeuAspThrGlnAlaIleAlaIyLeuThrArgHisIleAl 1275  
3573 GACCGCGCTGTTGCCGAGACCGCGCAACGCGTTTGACAAACGNGCA 3622  
1275 aGlnLysSerArgAlaAspAlaGlnLysAsnSerValTrpMetSerAsn 1292  
3623 TCCGGNACACCAAAACACTACCGTTCCGCAAGATTTCCGCGCTACCGGCA 3672  
1292 hreGlyArgGlyArgAspIyAlaSerAlaGlnIyArgArgPheSerSer 1308  
3673 CAACCGACCTGCGCAATCGGATGCAAGAAACCTCGGACGCGG... 3720  
1309 LysArgThrGlnThrGlnIleGlyIleAspArgSerLeuSerGlnLysnme 1325  
3721 CGCGTCCGATCCTGTTTCCGCAACCGGACCGAAACACTTGACAG 3769  
1325 tGlnIleGlyAlaIleuThrIySerAspSerGlnHisIleThrPheAsp 1342  
3770 ACGGCATCGGCACTCGGACGCGTTCGCGAGCGCGCGCTTTCCGCGCA 3819  
1342 lAla... GlnGlyLysAsnThrPheValGlnIleAlaAsnLeuIyGlnLys 1357  
3820 TACGCGATCGG... AGTTGCAATCGGATGACAGACGCGCGGCTT 3866  
1358 TyTylLeuAsnAspAlaTrpIyValAlaGlyAspIleIyAlaGlySe 1374  
3867 TAGCAGCGGCAATCTNCAAGAGGATCGAGGCAAAATCGCGCGCGG 3916  
1374 rLeuArgSerArgLeuGlnThrGlnIuLysAlaAsnPheAsnArgHis 1391  
3917 TGTGCAATTACGCGATCAGGACAGATACCGCGCGGCTTCCGCGGATTC 3966



```

1391 eTlLeGlnThrGlyLeuThrLeuGlyAsnThrLeuLysIleAsnGlnPhe 1407
3967 GGCATGACGACCGTACATCGCGCAACGCGCATTTGTCACAAAGACGA 4016
1408 GlnIleValProSerAlaGlyIleArgTyrSerArgLeuSerSerAlaAs 1424
4017 TTACCGCTACGAAACGTCATATGCGCACCCCGGCTTGGCTTCACAC 4066
1424 pYrLysLeuGlyAspSerValLysSerMetAlaValLysT 1441
4067 GNTACCGNCGGCGCATTAAGCAGATTAATTCATCAACCGCGCACAC 4116
1441 hLLeuThrIleGlyLeuAspPheAlaTyrArgPheLys...ValGlyAsn 1456
4117 ATNTCCATCAGCCTTATTNAGCCTGTCCTATACCGATCGCGCTCGGG 4166
1457 LeuThrValLysProLeuLeuSerAlaAlaTyrPhe...AlaAsnTyrG 1472
4167 CAAAGTCCGACACGCGTCATATCGCGCATGCTGCTCAGATTTGCGCA 4216
1472 yLysGlyGlyValAsnValGlyLysSerPheAlaTyrLysAlaAsp 1489
4217 AACCCGCGATGCGGATGCGGCGTAACGCCGAATCAAGGTTTCACG 4266
1489 sngIleGlnIleTyrSerAlaGlyValAlaLeuLeuTyrArgAsnValThr 1505
4267 CTGTCGNTCCAGCTGCGCGCGCAAGGCCAAGTGAAGCGGACGACACA 4316
1506 LeuAsnValAsnGlySerIleThrLysGlyLysGlnLeuGlyLysGln 1522
4317 CAGCGGCGGCGATCAATTAAGCTACCGCTGCG 4347
1522 sserGlyGlnIleLysIleGlnIleArgPhe 1532

```

seq\_name: SwissProt\_40:IGA4\_HAEIN

```

seq_documentation_block:
ID IGA4_HAEIN STANDARD: PRT: 1849 AA.
AC P45386:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria: Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OA NCBI_taxid=127;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NHRT HK61:
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
influenzae type 1 immunoglobulin A1 proteases."
RL J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR: CLEAVES HOST IMMUNOGLOBULIN A
PRODUCING INTACT EC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
certain pro-1-xaa bonds in the hinge region. No small molecule
substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

DR EMBL: M67491; AAA24968.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR000710; IGA_S6.
DR Pfam: PF02395; IGA1; 1.
DR PRINTS: PR00921; IGASERPTASE.
KW Hydrolase; Serine protease; Transmembrane; zymogen; signal.
FT SIGNAL 1 25
FT CHAIN 26 1021
FT PROPEP 1022 1849
FT ACT_SITE 299 299
SQ SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AEA CRC64;

```

#### alignment\_scores:

Quality: 1475.00 Length: 1924  
Ratio: 1.517 Gaps: 64  
Percent Similarity: 50.520 Percent Identity: 25.364

#### alignment\_block:

US-09-303-518D-651 x IGA4\_HAEIN ..

Align seg 1/1 to: IGA4\_HAEIN from: 1 to: 1849

```

64 CGCTTCGCGCTGCTTACTATGACATATGCGTGCCTTCGGCATCTCTCC 113
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThr 21
114 CCAAGCTTGGGCGGACACACTTATTTCGCGCATCACTACCAATTCATC 163
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
21 oTyrThrGluAlaAlaLeuValAlaArgAspAspValAspTyrGlnIlePhe 38
164 GCGACTTTGCGCGAAATTAAGGCAAGTTTGCAGTCGGGCGGAAAGATTT 213
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
38 rGAspPheAlaGlnAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54
214 GAGGTNTACAAACAAAAAGGAGAGTGTGGCGCAATCAATGACAAAGC 263
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
55 GluValAlaArgAspLysLysAsnGlnSerIleGlySerAlaLeuProAsnG 71
264 C...CGATGATGATTTTCTGTGTGTGCG...CGTAACGGCGGTGCGG 307
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
71 yLleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88
308 CATTGCTGGCGGATCAATATATATGAGCGTGCACAT...AACGGCGC 354
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
88 hLLeuValAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104
355 TATTAACAAGCTTGATTTTGGT.....GCGAAGGAGAGNAA 369
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
105 ValSerGlnLeuIleHisPheGlyAsnLeuAsnGlyAsnMetAsnGlnG 121
390 TCCGATCAGCAGCGT.....TTTCTTACCAAAATG 421
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
121 hAlaLysSerHisArgAspValSerSerGluGluAsnArgTyrThrY 138
422 TGAAGAATAATTAATTAAGCTGCACAT..... 450
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
138 alGluLysAsnAsnPheProThrGluAsnValThrSerPheThrThrLys 154
451 .....TCACACCTTACACAGCGGATTAATATGCGCGGTTT 488
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
155 GluGluGlnAspAlaGlnLysArgArgGluAspTyrTyrMetProArg 171
489 GCATAAATTTGTACAGATGAGAACTGTGCGAA...ATGACGAGTGACA 535
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
171 uAspLysPheValThrGluValAlaProIleGluAlaSerThrAlaAsn 188
536 TGAAGGGGAATACCTATTCGATTAAGAAAAATATCCGAGCGTGTCCGC 585
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

```

188 snasnlysglylutyryasnasnserAspLysTyrProAlaIleValArg 204  
586 ATGGCTGACGACACCACTAT.....TGGCGTAT..... 615  
205 LeuGlySerGlySerIleIleTyrLysGlySerArgTyrGlnIle 221  
616 .....GATGATGACAAACAGCGGATTTATCTACTCCGGCGAT 655  
221 uileLeuThrGlnLysAspLysGlnGlyLeuLeuArgasn.....T 236  
656 GATTATTTGGCGGCATACACATATCAGAGTTGGGAAATATATGCGCA 705  
236 rpsAlaGlyGlyAsp...AsnLeuGlnLeuValGlyAsnAlaTyrThr 251  
706 NTAGATTGAGCGGC.....GATGTGGCGCATGCGCAAGAC..... 741  
252 TyrGlyIleAlaGlyThrProTyrLysValAsnHisGlnAsnHisLys 268  
742 .....TATGCG..... 747  
268 uileGlyPheGlyAsnSerLysGlnHisSerAspProLysGlyIleL 285  
748 .....CCTATGCCGATTGACAGGTGCGGACGACAGCGGTG 786  
285 euSerGlnAspProLeuThrAsnTyrAlaValLeuGlyAspSerGlySer 301  
787 CCAATGTTTATTTATGACAAACAAACAATAATGCTGCTCAACGAGT 836  
302 ProLeuPheValTyrAspArgGlyLysTyrProLeuPheLeuGly.. 317  
837 TTTACAAACCGGCTACCTTATTCGGGAGGAAACGGTTTCCACGTGA 886  
318 .....SerTyrAspPheTyrAlaGlyTyrAsn..... 326  
887 TAGCAAAAGATTGTTCTACGATGACATTACAGAGCGCATACACTACC 936  
327 ..LysLysSerTyrGlnGlyTyrPasnIleTyrLysHisGlnPheAlaGlu 342  
937 GTCCTNTTTGAACCGCGCAGTACAGACATTTTCTTACATCCAAACA 986  
343 LysIleTyrGlnGlnTyrSerIleGly.....SerLeuThrGlySerAs 357  
987 C.....AACGGTACGGGTACGGTACAGAAA 1012  
357 nThrGlnTyrThrTyrPasnAlaThrGlySerThrSerThrIleThrGlyG 374  
1013 CCAACGAAAGGNTCCAAATCCAAAGCTTAAGTACAGACAGTCGACAG 1062  
374 LysIleGlnProLeu..... 378  
1063 TTTGACGAATCTTTGAATGAATGAATGAAGACAGTTTACGCGGACAG 1112  
379 .....SerValAspLeuThrAspGly..... 385  
1113 GGGTGTTAATCAGTACCGTCCAAAGTTAAACAACGGTGAACCTTTCTT 1162  
386 .....LysAspLysProAsnHisGlySerIleThrL 397  
1163 TTATCATTAACGCAACGCAACATCTTATCAAAACAACATCAACCAA 1212  
397 eu.....LysGlySerGlyThrLeuThrLeuAsnHisIleAspGln 411  
1213 GCGCGGCGGGTGTGATTTTGAAGTGATTTTACGGTC...TCCGCTGA 1259  
412 GlyAlaGlyGlyLeuPhePheGlnGlyAspTyrGlnValLysGlyThrSe 428  
1260 AAACAACGAACGTCGCAAGCGCGCGCTTCATATCAGTGAAGACAGTA 1309  
428 rAspSerThrThrTyrLysGlyAlaGlyValSerValAlaAspGlyLysT 445  
1310 CCGTTCCTTGGAAGTAACGCGCGTGGCAACGACCGCTGTCCAAATTC 1359  
445 hValThrThrLysValHisAsnProLysTyrAspArgLeuAlaLysIle 461  
1360 GGCAAGGACAGCTGACAGTTCAAGCCAAAGGGGAAACCAAGGCTGAT 1409  
462 GlyLysGlyThrLeuValValGlnGlyLysGlyLysAsnGlnLysLeu 478  
1410 CAGCGTGGCGGAGGTACAGTCACTTTTGGATCAGCAGCAGACAGTAAG 1459  
478 uLysValGlyAspGlyThrValIleLeuLysGlnLysAlaAspAlaAsn 495  
1460 GCAAAAACAGCCTTATGTAATTCGCTTGNTCAGCAGCGAGCGGTACG 1509  
495 snLysValGlnAlaPheSerGlnValGlyIleValSerGlyArgSerThr 511  
1510 GTGCACATGCAATGCCATTAATCACTTCACCCGCAACACTCTATTGCG 1559  
512 LeuValLeuAsnAspAspLysGlnValAspProAsnSerIleTyrPheGln 528  
1560 CTTTCGCGCGGACGCTTGGATTTAAAGGGCAATTCCTTCTTCCACC 1609  
528 yPheArgGlyLysArgLeuAspLeuAsnGlyAsnSerLeuThrPheAspH 545  
1610 GTATTCAAAATACCGATGCAAGGCGCATGATGNCMTCAATATGCCACA 1659  
545 IStIleArgAsnIleAspAspGlyAlaArgValValAsnHisAsnMetThr 561  
1660 ACAACATCCACCGTTTACCATTAACAGGAATGAAGTATTAACACACCG.. 1707  
562 AsnThrSerAsnIleThrIleThrGlyGlnSerLeuLeuThrAsnProAs 578  
1707 ..... 1707  
578 nThrIleThrSerTyrAsnIleGlnAlaGlnAspAspHisProLeuA 595  
1707 ..... 1707  
595 rGlnIleArgSerIleProTyrArgGlnLeuTyrPheAsnGlnAspAsnArg 611  
1707 ..... 1707  
612 SerTyrTyrThrLeuLysLysGlyAlaSerThrArgSerGlnLeuProGln 628  
1708 ...AGTGGTAAAGAT..... 1719  
628 nasnSerGlyLysSerAsnGlnAsnTyrLeuTyrMetGlyArgThrSerA 645  
1720 .....ATCAATGACTTATTTACAGCAAGAAATTT 1749  
645 spGlnAlaLysArgAsnValMetAsnHisIleAsnAsnGlnAlaArgMetAsn 661  
1750 GCCTACAAACGGTGGTTTGGCGAGAAAGATPACGACCAAAACG...AACGG 1796  
662 GlyPheAsnGlyTyrPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 678  
1797 GCGGCTCAACCTTGTTTACACGCCCGCGCAGAAAGACCGACCCGCTGC 1846  
678 LysLeuAsnValThrPheAsnGlnLysSerAspGlnAsnArgPheLeu 695  
1847 TTTCGCGGCGGACAAATTTAAAGCGCAACATCAGCAACAAACGCGAAA 1896  
695 euThrGlnGlyThrAsnLeuAsnGlnLysPheLeuAsnValGlnLysGlyThr 711  
1897 CTGTTTTTTCAGCGCGCAGACGACGACCGCGCCCTACATCAATTTAGAG 1946  
712 LeuPheLeuSerGlyArgProThrProHisAlaArgAspIleAlaGlyI 728  
1947 CGGGTGTCAAAATATGAAAGT.....ATCCCAAGAGGAAATCTGT 1990  
728 eSerSerThrLysLysAspProHisPheThrGlnAsnAsnGlnValVal 745  
1991 GGGACAACGACTGATCACCACGCTTTAAAGCGGAAATTTCCATATT 2040  
745 alGlnAspAspTyrPheAsnArgAsnPheLysAlaThrThrMetAsnVal 761



```

1327 ProGlnAlaGluProAlaArgGluAsnValLeuThrThrLysAsnValGln 1343
3445 ....CCGCAACCGCAGCCCAACCGCA..... 3468
||||| |||||
1343 yGluProGlnProGlnAlaGlnProGlnThrGlnSerThrAlaValProT 1360
3468 ..... 3468
1360 hrThrGlyGluThrAlaAlaAsnSerLysProAlaAlaLysProGlnAla 1376
3469 .....CCGCAACCGCAGCCGAGCGGCTGATNAG..... 3501
||||| |||||
1377 GlnAlaLysProGlnThrGluProAlaArgGluAsnValSerThrValAs 1393
3501 ..... 3501
1393 nrThrLysGluProGlnSerGlnThrSerAlaThrValSerThrGluGlnP 1410
3501 ..... 3501
1410 roAlaLysGluThrSerSerAsnValGluGlnProAlaProGluAsnSer 1426
3501 ..... 3501
1427 IleAsnThrGlySerAlaThrThrMetThrGluThrAlaGluLysSerAs 1443
3501 ..... 3501
1443 pLysProGlnMetGluThrValThrGluAsnAspArgGlnProGluAlaA 1460
3501 ..... 3501
1460 snThrValAlaAspAsnSerValAlaAsnAsnSerGluSerSerGluSer 1476
3502 .....CGTATGCAATACGGTTGAGTGAATTTCCGCCAC 3539
||||| |||||
1477 LysSerArgArgArgSerValSerGlnProLysGluThrSerAlaGln 1493
3540 GCTCAACAGCGTTTCCGCGTACAGAC..... 3567
||||| |||||
1493 uGluThrThrValAlaSerThrGlnGluThrThrValAspAsnSerValS 1510
3567 ..... 3567
1510 erThrProLysProArgSerArgThrArgArgSerValGlnThrAsn 1526
3568 .....GAATTGACCGCGCTTTGCCGAGACCGCGCA 3602
||||| |||||
1527 SerTyrGluProValGlnLeuProThrGluAsnAlaGluAsnAlaGluAs 1543
3603 CGCN..... 3606
||||| |||||
1543 nValGlnSerGlyAsnAsnValAlaAsnSerGlnProAlaLeuArgAsnL 1560
3606 ..... 3606
1560 euThrSerLysAsnThrAsnAlaValLeuSerAsnAlaMetAlaLysAla 1576
3606 ..... 3606
1577 GlnPheValAlaLeuAsnValGlyLysAlaValSerGlnHisIleSerGln 1593
3607 .....GTTTGACACAGCNCATCC 3625
||||| |||||
1593 nLeuGluMetAsnAsnGluGlyGlnTyrAsnValIrrPILeserAsnThrS 1610
3626 GGNACACCAACACTACCGTTGCAAGATTTCGGCGCTACCGCAACAA 3675
||||| |||||
1610 erMetAsnLysAsnTyrSerSerGlnTyrArgArgPheSerSerLys 1626
3676 ACCGACCTCGGCAATCGTATGCAGAAACCTCGGAGGCG...CG 3722
||||| |||||
1627 SerThrGlnThrGlnLeuGlyTyrPaspGlnThrIleSerAsnAsnValGln 1643

```

```

3723 CGTGGCATCTCTTTTCGCACACCGGACCGAAACACTTCGACGACG 3772
||||| |||||
1643 nLeuGlyGlyValPheThrTyrValArgAsnSerAsnAsnPheAspLysA 1660
3773 GCATCGGCACACTCGGCACGCGTTCGCCAGCGCGCTTTTCGGCAATAC 3822
||||| |||||
1660 lAsrSerLysAsnThrLeuAlaGlnValAlaGlnPheTyrSerLysTyrTyr 1676
3823 GGCATCGGCAGCTTCGACATCGCATCGACGCGCGCGGTTTATACAG 3872
||||| |||||
1677 AlaAspAsnIleStrpTyrLeuGlyIleAspLeuGlyTyrGlyLysPheGln 1693
3873 CGGCATCTTTCACAGCAGCGCATCGGAGCAAAATCGCGCGCGCTGTC 3922
||||| |||||
1693 nSerAsnLeuGlnThrAsnAsnAlaLysPheAlaArgHisThrAlaG 1710
3923 ATTACGGCATTCAGGACGACGATACCGCGCGTTTCGGCGGATTCGGCATC 3972
||||| |||||
1710 nIleGlyLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnPheAlaVal 1726
3973 GACCGTACATCGCGCAGCGCGCTATTTCGTCACAAACGATTCACG 4022
||||| |||||
1727 LysProThrValGlyValArgTyrSerTyrLeuSerAsnAlaAspPheAl 1743
4023 CTACGAAACGTCATATCGCCACCGCGCTTCGCTGACCGCATAC 4072
||||| |||||
1743 aLeuAlaGlnAspArgIleLysValAsnProIleSerValLysThrAlaP 1760
4073 GNGCGGCATTAAAGCAGATTCATTCAACCGCGGCAACACATTC 4122
||||| |||||
1760 nAlaGlnValAspLeuSerTyrThrTyrHisLeuGlyGlu...PheSer 1775
4123 ATCAGCCTTATTNAGCCTCTCTATACCGAGCCGCTTCGGGCAAGT 4172
||||| |||||
1776 IleThrProIleLeuSerAlaArgTyr...AspAlaAsnGlnGlyAsnGln 1791
4173 CCGAACAACGTCATATCGCCACCGCGTATTCGCTGAGATTCGCGCAACCC 4222
||||| |||||
1791 yLysIleAsnValSerValTyrAspPheAlaTyrAsnValGluAsnGln 1808
4223 GCAGTGGGGAATGGGCGTAACCGCGCAATCAAAAGTTTCAGCGTCC 4272
||||| |||||
1808 nGlnTyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSer 1824
4273 NTCACGCTGCGCGCGCAAGCGCAACCTGGAACGCAACACGCGC 4322
||||| |||||
1825 LeuIleGlyLeuThrLysAlaLysGlnAlaGlnLysGlnLysThrAl 1841
4323 GGCATCAATTAAGCTAC 4341
||||| |||||
1841 aGluValLysLeuSerPhe 1847
seq_name: SwissProt_40:YPIA_ECOLI
seq_documentation_block:
ID YPIA_ECOLI STANDARD; PRT: 1569 AA.
AC P52143; P76610; P77017; P77019;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical outer membrane protein yPjA.
GN yPjA OR B2647.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=9742617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

```



480 TyrThrGlyGly.....IleAlaHisGlyValAlaSerGlnGluThrGln 493  
1221 CCGTTTGTATTTTGAAGGTGATTTTACGGTCTCGCTGAAACAAACGAAA 1270  
::: :  
493 ySerAlaLeuValAlaAsn..... 499  
1271 CGTGGCAAGCGCGCGCTTCATATAGT.....GAAAGC 1305  
||||| :  
500 ....ThrGlyAlaGlyThrAspIleGluGlyTyrAsnLysSerHis 514  
::: :  
1306 AGTACGTTACTCTGAAAGTAAACGGCGGCGGCAAGCAGCCGCTGCA 1355  
||||| :  
515 PheThrIleThr.....GlyGlyGluAlaAsn..... 523  
1356 AATCGCAAAAGCAGCGTCGACGCTTCACGCAAAAGGCAAAACGAGCT 1405  
::: :  
524 .....TyrValValLeuGluAsnThrGlyG 532  
1406 CGATCAGCGTGGCGGCGGCTACGTCATTTTGATCAGCAGCAGCAGT 1455  
::: :  
532 IuThrValValAlaLysThrSerAlaLysAsnThrThrIleAspThr 548  
1456 AAGGCAAA.....AACAAGCCTTAGTGAATCGCTTGN 1493  
||||| :  
549 GlyGlyLysLeuIleValGlnLysGluAlaLysThrAsp..... 561  
1494 CAGCGCAGGGGCTACGCTCAACTGAATCGCATATACAGTTCAACCCG 1543  
::: :  
562 .....SerThrArgLeuAsn..... 566  
1544 ACAACTGTATTTGGCTTCGGCGGCGAGCTTTGGATTAAACGGCGAT 1593  
::: :  
567 .....AsnGlyGlyValLeuGluValGlnAspGly 576  
1594 TCGCTTGTCCACCGTATTCAAATACCGATGAAGGGCGCATATG 1643  
::: :  
577 GlyGluAlaLysHis...ValGlnGlnSerGlyGlyAlaLeuIle... 591  
1644 CNACTAATATGCCACAAACATCCACCGCTTACATTCACGGAATGAAA 1693  
||||| :  
592 .....AlaSerThrThrSerGlyThrLeuIleGluGlyThrAsn 605  
1694 GT.....ATTACAAACGAGTGGTGAATATCAAT 1725  
|| :  
605 eTyrGlyAspAlaPheTyrIleArgAsnSerGluAlaLysAsnValAl 621  
1726 AGCTTAATTTACAGCAAGAAATGGCTACAGGTTGGTTGGGAGAGA 1775  
|| :  
622 LeuGluAsnAlaGlySerLeuThrValValThrGlySerArgAlaValAs 638  
1776 AGATACGACCAAAACGACGCGGCTCAACCT..... 1809  
638 pThrIleIleAsnAlaAsnGlyLysMetAspValTyrGlyLysAspValG 655  
1810 .....GTTTACCAGCCCGCCGA 1827  
655 IYThrValLeuAsnSerAlaGlyThrGlnThrIleTyrAlaSerAlaThr 671  
1828 GAAGACCGCAGCCNGCTGCTTCGCGGCAACAAATTTAAAGGACACAT 1877  
672 SerAspLysAlaAsnIleLysGlyLysGlnThrValTyrGlyLeuAl 688  
1878 CACGCAAAACAAAC.....GGCAAACTGTTTTCAGCGGCAACCGCA 1918  
688 aThrGluAlaAsnIleGluSerGlyGlnIleValAlaSpGlyGlySerT 705  
1919 CACGCGACCCCTACATCATTTAGAGAGCGGCTGCAAAAATGAGAGT 1968  
705 hGlyLys.....ThrHisIleAsnGlyGlyThrGlnThrValGlnAsn 719  
1969 ATCCCAACAAGGA.....GAATTCGTGTGGCAACGAGCTGATCNA 2009  
720 TyrGlyLysAlaIleAsnThrAspIleValSerGlyLeuGlnIleLeu 736  
2010 CCGCAGTTTAAACCGGAAATTTCCATATTCAGCGCGGCGCGCTGA 2059  
::: :  
736 tAlaAsnGlyThrAlaGluGlySerIleIleAsnGlyGlySerGlnVal 753  
2060 TTCCCGC..... 2067  
753 aAlaSerGluGlyGlyLeuAlaGluAsnSerValLeuAsnAspGlyGlyThr 769  
2068 ...AATGTGCCAAGTGAAGGCGATTGNCATTTGAGCAATCGCGCA 2114  
770 LeuAspValArgGluLysGlySerAlaThrGlyIleGlnGlnSerG 786  
2115 AGCAGTTTGTGTCGACCGCAGCAGCAAGCAGATATGTCAGCTT 2164  
786 nGlyAlaLeuValAlaThrThrArgAlaThrArgValThrGlyThrArg 803  
2165 CGACCTGACN.....GCTCGCAAT..... 2187  
803 IaAspGlyValAlaPheSerIleGluGlnGlyAlaAlaAsnIleLeu 819  
2188 .....TGTGTGAAANCAATTCACGCGATA 2216  
820 LeuAlaAsnGlyGlyValLeuThrValGluSerAspThrSerAsp 836  
2217 AGTATTCCTTCATTG.....ACTAAGCAGCAGC 2245  
836 sThrGlnValAsnMetGlyArgGluIleValLysThrLysAlaThr 853  
2246 TNAAGCGCANTGTNAGCTNCCNATNAGTNTTAAANCTCNCNGG 2295  
853 IaThrGly.....ThrThrLeuThrGlyGly 861  
2296 CNTGCNMCACITNANGC..... 2313  
862 GluGlnIleValGluGlyValAlaAsnGluThrThrIleAsnAspGly 878  
2314 .....AATCTAGTCAATGCGCAT.....ACAGTTATPACG 2347  
878 yIleGlnThrValSerAlaAsnGlyGluAlaIleLysThrLysIleAsn 895  
2348 TCAGCCCAACGCCCAACGCGCAAC.....CTTACCTCGTG 2388  
895 IuGlyGlyThrLeuThrValAlaAsnAspAsnGlyLysAlaThrAspIleVal 911  
2389 GGCATGCCCAAGCAATTTAATCAGCCACATTTAAGGCAACNATC 2438  
912 GlnAsnSerGlyAlaAlaLeuGlnThrSerThrAlaAsnGly...IleG 927  
2439 GGNFTCGGCGCAAT.....GCTTCATTTAATCTPACGACAGCGCG 2479  
927 uIleSerGlyThrHisGlnIleGlyThrPheSerIleSerGlyAsnLeu 944  
2480 CACAAC.....GGCAGTGTGACGCTTTCGACAC 2511  
944 IaThrAsnMetLeuLeuGluAsnGlyGlyAsnLeuValLeuAlaGly 960  
2512 GCTAAGCAAAAGTAAAGCAATTCGCGACCTCAAGGCAATTCGCCAGC 2561  
961 ThrGluAlaArgAspSerThrValGlyLysGlyGly... 972  
2562 CGATPAGCAGATATTCATTTTGAACACAGCCCTTACCGCAACTCA 2611  
973 .....AlaMetGlnAsnLeuGlyGlnAspSerAlaThrLysValAsn 987  
2612 GCGCGACCAAGANACAGCATTTACACTTAAAGACGAGATGACGCTG 2661  
987 eTyrGlyGlnGlyTyrThrLeuGlyArgSerLysAspGluPheGlnAlaLeu 1003  
2662 CCGTCAGCGACGGAATTA..... 2679  
1004 AlaArgAlaGluAspLeuGlnValAlaGlyGlyThrAlaIleValTyrAl 1020

```

2680 .....GCAATTAAAC 2692
1020 aglythrleualaservalserglyalathrcllyserleuserl 1037
2693 TT..... 2694
1037 eumethrproarspaspaValThrProValLysleugllyalaval 1053
2695 .....GACACGCCACCTTACCTCAATCCGCTAT..... 2727
1054 ArgilethraspseralathrleuThleugllysnilyalaspThrTh 1070
2728 ..CGCCAGATGCTGCAGCGCGCAACCGCAGNCTG..... 2763
1070 rleualaspLeuthrAlaIaservArglySerValThrleuasnsera 1087
2764 .....TCAGACAGCGCGCGCGCTGCGGC...CGTCCCTA 2799
1087 snasnsercysalaglYThrSerascysglutYrArgvalasnserleu 1103
2800 TTA.....TCGGTACACCGCCGAC 2819
1104 leuleuasnapglyaspvallyrleuseralaglInthrAlaIalaproal 1120
2820 TTCGGTAGAATCCGCTTCAACACCGCTGACGGTAACCGCAATTAACN 2869
1120 ethrthrasnnglyleYrAsnthrleuthrthrasn...Glueuserg 1136
2870 GTCAAGAACATTCGCTTTATGTGGAACCTTCTGGCTACCGAAGCAGC 2919
1136 lysercylasnpheYrleuHsrthrasnvalalaglYserArglyasp 1152
2920 AATTGAAGCTGGCGAAAGTCCGAAAGNCTTACACTGGCGGCGTAA 2969
1153 GluevalvalasnasnasnalathrcllysnphelysIlephevalgl 1169
2970 CAATACCGCAACGACCGGTAACCTCATCAATTTGACGTAAGTGAAG 3019
1169 naspthrlyValserProglInserAspaspalameThrleuVal.... 1184
3020 GGAAGACAAACAACCGCTGTCCGAAACCTTAATTTCACCCCTGTAAC 3069
1185 .....LysThrlygllyglYaspAlaSerPheThrleuLysn 1197
3070 GAA.....CACGTGATCGCGCGCGCTGACCAATCATCGGAA 3113
1198 ThrlygllyphevalaspLeuglYThrlyglYr...ValleuLysSe 1213
3114 AGACGAGC.....GAGTCCGCTCATATCCGCTCAAGAACAAAGAGC 3157
1213 raspglyasnserasntrPasnleuthrasnspVallys..... 1226
3158 TTTCGCAACAACCTCGGCAAGCAGAACCCAAAAACAGCGGAAAAAGAC 3207
1226 ..... 1226
3208 AACGGCAAAAGCCTTGACGCGCTGATTTGGCGCGCGCGGATCGCCGCA 3257
1227 .....Pr 1227
3258 AAAGACAGAAAGCTTGGCAACCGCGCGCGCGAGCGGCAAAATG 3307
1227 oasnProaspProIleProasnPro..... 1235
3308 TCGGCAATTATGACGCGGAGAGAAAGAAAAACGGGTGACGCGGATAA 3357
1235 ..... 1235
3358 GACAGCGCNTTGGCAAAACGCGGAGAAACCGCGCGGNTACAC 3407
1236 .....LysProaspProLysProaspProly 1244
3408 CCGCTTCCCGCGCGCGCGCGCGCGGATTTGCCGCAACCGCAGC 3457
1244 sproaspProasnProlys.....ProaspProThrp 1255
3458 CCCAACCGCAACCTCAACCCCAACCGCGCGCGACTGATNAGCCTTAT 3507
1255 roaspProThrpProthProvalProglulysArglyleThrPro..... 1269
3508 GCCAATGCGGTTTGAGTAATTTCCGCCACCTCAACGAGCTTTTC... 3555
1270 SerthralalavalLeuasnmetAlaIalThrleuProleuValpheas 1286
3556 .....GCCGTACAGACGAATTGACCGCGGTGTTGCCGAA 3592
1286 palagluleuasnserrlearggluargleuasnlleuMetYalaserp 1303
3593 ACCGCCCAACGCGNCTTGGACAAGCAGCATCCGGAACCAACCAACTAC 3642
1303 roHIsAsnasnasvalTrpGlyAlaThrYrAsnthrArgasnasVal 1319
3643 CGTTCGCAAGATTTCGCGGCTACCGGCACAAACGACCTGGCGCAAT 3692
1320 ThrThraspalaglYalaglYpneuglInThrleuthrclYmethrva 1336
3693 CGGTATGCAAGAAAAAC.....CTCGCAGCGCGCGCGCGCA 3730
1336 lglYleaspserrargnasnpIleprogluglyleThrThleuglyA 1353
3731 TCCGTGTTTGGCACACCGGACCGAAACNCTTCGACGACGCGATCGGC 3780
1353 lapmetglYrYserHIsSerHIsleaglYpheaspArgly..... 1367
3781 AACTCGCACGCGCTGGCCACGCGCGCTTTCGCAATACGCAATCGG 3830
1368 .....GlyHIsglySerVal...GlyserYrserleuGl 1378
3831 C..... 3831
1378 yglYThrAlaserTrpLuhIsgluserglYpethYrleuaspglyValY 1395
3832 .....AGSTGACATCGGCATC.....ACACGCGC 3858
1395 alYsleuasnasnargphelysSerasnvalalaglYlYmetSerSergly 1411
3859 GCGGGTTTGAAGCGCGCANTCTTTCAGACGCGATCGAGCAAAATCCG 3908
1412 GlYalalasnnglyserYrHIsSerAsnlyleuclYlYHIs..... 1426
3909 CCGCGCGGTGCGATTACGCGCATTCAGGACGATACCGCGGTTTCG 3958
1427 .....IlegluThrlyglYmetArgPheThrAspg 1436
3959 GCGGATTCGCGATCGAACCGTACATCGCGCGCAACGCGCTATTTCGCA 4008
1436 lyasntrPasnleuthrProlyThrAlaserleuthrPheThrAlaasp 1452
4009 AAAGCGATTACCGCTACGAAAC.....GTCAATATCGGCAC 4046
1453 AsnProglulYrHIsleuserAsnlyMetYlYserlYsSerValaspth 1469
4047 CCGCGGTCTTCGCTTCAACCGNTACGN...GCGGCGATTAAAGCAGATT 4093
1469 rArgSerIle.....TyArggluleuglyAlaThrleuserl 1482
4094 ATTCAATCAACCGCGCAACACATNTTCATACNCTTATTNAGCTG 4143
1482 yrasmetargleuglyAsnlyMetgluValgluProThrleuLysAla 1498
4144 TCCATATACGATGCGGCTTCGGCAAGTCCGAAACGCGGTCAATACCGC 4193
1499 AlavalArglysluphevalaspaspasnarvalYlYvalasnseras 1515
4194 NGTATGCTCAGGATTTC.....GGCAAAACCGCGATGCGGAATGGG 4237

```



```

1515 pGlyAsnPhenValAsnTyrLeuSerGlyArgArgGlyIleTyrGlnAlaG 1532
4238 GCCTAAACGCCGAATCAAGTTTCACGCTGTCCATCCAGCGTCCGCC 4287
1532 LyleIySalaserPheSer...ThleuSerGlyHisLeuGlyVal 1547
4288 GCCAAGGCGCACTGCAAGCGCAACAGCGCGCATCAAA..... 4332
1548 Gly.....TyrSerHisSerAlaGlyValGluSerPr 1558
4333 .....TTAGCTACCGCTGG 4347
1558 oTPAsnAlaValAlaGlyValAsnTrp 1567
seq_name: SwissProt_40:BIGA_SALTY

seq_documentation_block:
ID BIGA_SALTY STANDARD; PRT; 1953 AA.
AC P25927; P25928; Q9XCQ3;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Putative surface-exposed virulence protein biga precursor.
GN BIGA OR STM3478.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=14028;
RA Stojiljkovic I., Valentine P., Heffron F.;
RT "Salmonella typhimurium rbs homolog.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Lt2 / SGSC1412 / ATCC 700720;
RA McCelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT Lt2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE OF 1-765 FROM N.A.
RC STRAIN=Lt2;
RA MEDLINE=91100301; PubMed=1987123;
RA Wu J.Y., Siegel L.M., Kredich N.M.;
RT "High-level expression of Escherichia coli NADPH-sulfite reductase:
RT requirement for a cloned cysG plasmid to overcome limiting sigmaHeme
RT cofactor";
RL J. Bacteriol. 173:325-333(1991).
RN [4]
RP CAUTION: Ref.3 sequence differs from that shown due to frameshifts
RN [5]
RP In positions 414 and 732.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; AF133686; AAD39458.1; -
DR EMBL; AE008859; AAL22340.1; -
DR EMBL; M64606; AAA27042.1; ALT_FRAME.
DR EMBL; M64606; AAA27043.1; ALT_FRAME.
DR PIR; C39200; C39200.
DR PIR; D39200; D39200.
DR StyGene; SG10437; BIGA.

```

```

KW Virulence; Repeat; signal; Complete proteome.
FT SIGNAL 27
FT CHAIN 28 1953
FT DOMAIN 101 252
FT REPEAT 101 103
FT REPEAT 104 113
FT REPEAT 114 122
FT REPEAT 123 133
FT REPEAT 134 144
FT REPEAT 145 155
FT REPEAT 156 166
FT REPEAT 167 177
FT REPEAT 178 188
FT REPEAT 189 199
FT REPEAT 200 210
FT REPEAT 211 221
FT REPEAT 222 232
FT REPEAT 233 243
FT REPEAT 244 252
FT CONFLICT 207 207
FT CONFLICT 514 514
FT CONFLICT 1698 1698
FT CONFLICT 1795 1798
FT CONFLICT 1836 1837
SQ SEQUENCE 1953 AA; 200150 MW; 611B3F1C954D91AE CRC64;

```

```

alignment_scores:
Quality: 264.50 Length: 1669
Ratio: 0.359 Gaps: 74
Percent Similarity: 44.158 Percent Identity: 18.394

```

alignment\_block:

US-09-303-518D-651 x BIGA\_SALTY ..

Align seg 1/1 to: BIGA\_SALTY from: 1 to: 1953

```

217 GTNNACACAAAGAGGAGTGTGCGCAATCAATGACAAAGCC 266
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
413 ValAspAsnAlaThrGlyAlaLeuIleSerGlyAsnGlyThrThrAs 429
267 GATGATTGATTTTCTGTGTCGCGTAAACGGCGGCATGCTG 315
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
429 nPheAlaGlyAspIleAlaValSerGlyGlyThrAlaIleIleAla 446
316 ..GGCGATCATATATGTGTGAGCGTGCACATACGGCGCTATAACA 363
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
446 spGlyAspAsnAlaThrIleLys.....AsnThrGlyThrSerAsp 459
364 GTTGAT...TTTGTGCGGAAGAGNAATCCGATCAGACGGTTTTC 410
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
460 IleSerGlyAlaGlySerThrGlyThr..... 468
411 TTACCAATTGTGAAAGAAATAATTAAGCTGACATTCACCCCTT 460
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
469 .....ValIleAspGlyAsnAsnAlaArgValAsn..... 479
461 ACAAGGGGATTAATCATATGCGCGTTTGCAATAATTGCACAGATGCA 510
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
480 ..AspGlyAsp.....MetThrIleThrAspGly 488
511 GAACCTGTCGAATATGACAGATGACATGAGGGGAACTATTCGGATA 560
488 ..... 488
561 AGAAATAATATCCGAGCGTGTCCGATCGCTCAGACACCACTATTGGC 610
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
489 .....GlyThrGlyGlyHisIle...T 495
611 GTTATGATGAT.....GACAAACAGCGCATTTATCTCTACTCGGC 651
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
495 hrGlyAspAsnValAlaIleAspAsnAlaGlySerThrThrValSerGly 511

```

652 GCA.....TGTTATTTGGCGGCAATACATATGACAGG 686  
|||.....:|:| | | | | : : : :  
512 AlaSpAlaThrAlaLeuTyrIleGluGlyAspAsnAlaLeuValIleAs 528  
687 TTGGGAAATATGCGCTAATTACT.....TTGAGCG 718  
: | | | | : : : : : : : : : : : : : :  
528 nglGlyAsnGlnThrIleSerGlyGlyAlaValGlyThrArgIleAsp 545  
719 GCGATGCGCGCATGCCAGACGATATGCCCTATGCCGATTCGAGTGGC 768  
||| | | | : : : : : : : : : : | | | | |  
545 LysAspAspAlaHisThrThrAsnThrGlyAspIleAlaValAspGlyAla 561  
769 GCA.....GGCGACAGCGGTCG..... 786  
: : : : : | | | | | : : : : :  
562 GlySerAlaAlaValIleIleAsnGlyAspAsnGlySerLeuThrGlnAl 578  
787 ....CCAATGTTTATTTATGACAAACAAACATTAATGGCTGCTCAACG 832  
: : : : : | | : : : : : : : : : :  
578 aglyAspLeuLeuValThrAspGlyAla.....MetG 589  
833 GAGTTTACAAACCGGCTACCTTATTCGCGACGAGAAACGTTTCCAG 882  
| | | | | : : : : : : : : : : : : : :  
589 LylleIleThrTyrGlyThrGlyAsnGluAlaLysAsnThrGlyAsnAla 605  
883 CTGATACGCAAGATGGTCTACAGATGACATTTACAGAGCGCATACACA 932  
606 ThrValAlaGspAlaAspSerValGlyPheValAlaAlaGlyGluLysAs 622  
933 TACCGCTCTTTTGAACCGCGCAGTACGACATTTTCTTCTTACATCCA 982  
622 nThrPhe.....LysAsnLysGlyAspIleAspValSerLeuA 635  
983 ACAACACGCTAGCGGTACGCTAACACAGAAACGAAAGCTTCCAT 1032  
| | | | | : : : : : : : : : : : : : :  
635 snGlyThrGlyAlaLeuValSerGlyAspMetSerGlnValThrLeuAsp 651  
1033 CCAAGCTTAAGTACAGACAGCTCGACCTTTGACAGATCTTGAATGA 1082  
652 GlyspIleAsnValValSerValGln..... 660  
1083 AACTGATTAAGACCACTTTAC...GGCGACGGGGGTCTT..... 1119  
| | | | | : : : : : | | | | | : : : : :  
661 ....AspSerGluGlyValPheSerSerAlaThrGlyValSerValSerG 676  
1120 ..AATCACTACCGTCCAAAGTTAAACAACGCTGAAAACCTTCTTTATC 1167  
:  
676 LysAspSerAsnAlaValAspIleThrGlyAsnValAsnIleSer...Ala 691  
1168 GATTACGCG..... 1176  
| | | | | : : : : : : : : : : : : : :  
692 AspyrGlyGlnAspAspLeuAlaAlaGlyAlaProProLeuThrGlyVal 708  
1177 .....ACGCGCAACCTCATCT 1192  
708 lValValGlyGlyAsnGlyAsnThrValThrLeuAsnGlyAlaLeuAsnI 725  
1193 TATCAACACATCAACCAAGCGCGGGCTTTGATTTTGA...GGT 1239  
: : : : : | | | | | : : : : : | | | | | : : : : :  
725 LeaspAspAsnAspLeuSerAlaThrGlyGlyIntYrLeuAspValVal 741  
1240 GATTTACGCTCGCTCGGTAAGAAACAAGAAACGTGCCAAGCGGGGCGCT 1289  
: : : : : | | | | | : : : : : : : : : : : : : :  
742 GlyLeuSerValThrGlyAspAsnAspValGluIleAspGlyGlyI 758  
1290 TCATATC.....AGTGAAGAC..... 1305  
: : : : : | | | | | : : : : : : : : : : : : : :  
758 eAsnIleThrHisSerGluAspProLeuAspGlyThrSerAlaSpliet 775  
1306 .....AGTACCGTTACTTGAAAGTAAAGCGC 1332  
775 hrGlyIleSerValSerGlyAsnSerThrValThr.....LeuAsnGly 789

1333 GTGGCAACGACCGCCTGTCCAAATCGGCAAGCGACGCTGCAGTTCA 1382  
: : : : : : : : : : | | : : : | |  
790 HisSerThrIleAspThrAsnThrValAlaGlyGlyHisVal...Valle 805  
1383 AGCCAAAGGGAAGAAACCAAGGCTCGATCAGCGTGGCGACGCTACAGTCA 1432  
| | | | : : : | | | | : : : : : : : : : : : : : :  
805 uAlaArgValAsnAsnGlyGlySerLeuIleGluGlyAspAspSerValV 822  
1433 TT..... 1434  
:  
822 alaSpValAsnValSerTyrIleProThrGlyTyrTyrThrTyrAsnAla 838  
1435 .....TTGATCAGACGACAGAC... 1452  
:  
839 LeuLeuMetAlaAspGlyGluGlyThrSerIleGluAsnLysGlyAspI 855  
1453 .....GATAAAGCAAAANAC 1468  
855 eThrSerHisGlyValTyrSerValIleArgAlaAspAsnGlySerGluV 872  
1469 AACCTTTAGT..... 1479  
: : : | | : : : : : : : : : : : : : :  
872 alSerAsnSerGlyAspIleLeuValTyrAlaThrSerSerAsnSerSer 888  
1480 .....GAAATCGGCTTGNTCAGCGCAGCGGTACGGTGCACCT 1517  
| | : : : | | | | | : : : : : : : : : : : : : :  
889 GluAspArgAlaAlaIleThrArgAlaSerGlyGlySerAlaValHis 905  
1518 GAAT.....GCCGATATCATGTTCAACCCGACAACTCATTTGCG 1558  
| | | : : : | | : : : : : : : : : : : : : :  
905 sAsnLysAlaGlyGlyAspIleThrLeuIleSerAspGlnThrProGlnG 922  
1559 GCTTGGCGCGGCGCGCTTG...GATTAAACGCGCATTCGCTTCGTTG 1605  
| | : : : | | : : : | | : : : : : : : : : : : : : :  
922 LysSerGlyGlyIleGluValTyrProLeuLysTyrPyrThrHisThrPhe 938  
1606 CACCGTATTCAAATACGATGAGGAGGCGGATGATGNCATCATATATGCG 1655  
: : : : : : : : : | | : : : : : : : : : : : : : :  
939 TyrAlaMetMetAlaSerAspTyrGlyAspValAlaAsnAspGlyVal 955  
1656 CACAACAATCCACCGTTACCATTTACAGGAAATGAAGTATTTACACAAC 1705  
| | | : : : : : : : : : | | : : : : : : : : : : : : : :  
955 aThrIle.....HisLeuGlnGlyAlaGlyValTyrGlyValThrAlas 970  
1706 CGAGTGTGAAGATATCATATAGCTTAATTACGCAAAACAAATGGCTTAC 1755  
| | | | : : : | | : : : : : : : : : : : : : :  
970 eArgGlyLysAlaLeuAsnGluGlyAsnIleTyrIleuAspGlyLeuVal 986  
1756 AACGTTGGTTTGGGAGAAAGTACGACCAACAAACGAGCGGGGCTCAA 1805  
987 ProThrLeuAspAspGluAsnAsnIleThrSerThrSer..... 999  
1806 CTTTGTTTTACACCGCGCGCAGAACGCGCACCG.....CTGCTTT 1849  
: : : | | | | : : : : : | | | | : : : : : : : : : :  
1000 ....TyrTrpGlnProSerSerLeuTyrLeuThrSerSerGlyMetValA 1015  
1850 CCGCGGAGACAAATTTAAACGCCACATCAACGCAAAACAGGCAACTG 1899  
: : | | : : : | | : : : : : | | : : : : : : : : : : : : : :  
1015 lAdGlySerThrAspAlaAspGlyAspAlaThrAlaIleAsn..... 1028  
1900 TTTTTCAGCGCGACCGACCGCAGCGCTCAATCATTTAGGAAGCGG 1949  
: : : | | : : : | | : : : : : | | : : : : : : : : : :  
1029 .....ThrGlyAsnIleThr.....ValAsnAsnAlaGlyPheG 1040  
1950 GTGGTCAAAAATGGAAGT.....ATCCACAAGAGGAATGTGT 1990  
| : : : : | | : : : | | : : : : : | | : : : : : : : : : :  
1040 yMetMetAlaLeuAsnGlyGlyThrAlaIleAsnGlnGlyValIleThrL 1057  
1991 GGCACAACGAC..... 2001  
1057 euthrAlaAspAspGlyValThrGlyGlnAlaAspGluLeuValGlyMet 1073  
2002 .....TGCATCNCAGCAGCTTTAAAGCGGAA 2030

```
1074 AlaAlaLeuasnlglyValValIleasnAspThrSerglyValIleAs 1090
1090 nleasnAlaAspThrSerglyValIleAsnleuAspThrSerglyValIleAs 1107
1090 nleasnAlaAspThrSerglyValIleAsnleuAspThrSerglyValIleAs 1107
2081 TCGAAGCGCATTTGCAATTTGCAATTCACGCCCAAGAGTTTTCGTC 2130
1107 lelleasnAsnlglySerIleAsnleuAsnlglySerProMetAspAspThr 1123
2131 GCACCGCATCAAGCCATCATCTGTACAGCTTCGGACTGACNGGCTCT 2180
1124 AspSerHisMetGlyGlyThr..... 1130
2181 GACAATTTGTGCAANAANCATTACCGCATTAAGTATTGCTCAT 2230
1131 .....ProThrAspIleIleThrIleGlnSerL 1140
2231 TG.....ACTAAGACNGACNTNAGCGGACANTGTN 2259
1140 euproGlySerGlyAspSerAspThrArgThrSerAspThrGly..... 1154
2260 AAGCTTNCNATNAGCNTNTNTNAANGCNCNGGCGTGCNMCCTNAA 2309
1155 .....PhePheThrAlaGlyThrLeuAlaAsnThrGlyThrLeuAs 1170
2310 NSGCAATCTAGTGCATAATGCGATACAGTATACA...GTCAGCCACA 2356
1170 nGlyAspValAspValAsnGlyGlyThrLeuThrAsnGlyAlaAs 1187
2357 AGCCACCCAAAGCGCAACCTTACCTGCTGGGCAATGCCCAACA 2406
1187 erleuThrValAsnGlyThrValThrIleAsnGlyGlyAlaAsnAlaLeu 1203
2407 TTTAATCAAGCCATTAAGCGCAACNATCGNTTCGGCATGCTC 2456
1204 AlaAsnThrGlyThrLeuAspAlaAspAlaIleSerThrPheSerLe 1220
2457 ATTAACTCAAGCAACAGCGCGCAAAAAGCGAGTCTGACGCTTCG 2506
1220 upheAsnGlyAlaAspGlySerIleThrThrAspLeuThr..... 1234
2507 ACAACGCTTAAGCAACGTAAAGCCATTCGCGACGACGCGAGTCTC 2556
1235 .....LeuAsnGlyAspValThr 1240
2557 CTAGCCGATAAGCAGTATTCATTTGAAAAACGCCGCTTACCGGACA 2606
1241 Phe.....TyrAsnAsnGlyAspPheThrGlySe 1250
2607 ACTCAGCGCAGACAGANAGCATTAACCTTAAGACAGACGANTGA 2656
1250 rIleAlaGlyThrSerThrGlnGlnGlnIleValAsnThrGlyAspMet 1267
2657 CGCTG...CCGTCAGCGACGGAATTAGCAAT.....TTA 2688
1267 hrValAlaGlyAspGlySerLeuValSerGlySerPheThrPheThr 1283
2689 AACCTTGACAGCGACCATTAACCTTAACCTTCGCGCATTCGCGAGATC 2738
1284 AsnGlyGlyAspAlaThrLeuThr.....AspSerGly 1294
2739 TGCAGCGCGCAACCGCGCAGNGTGTACACAGCGCGCGCGCTTCG 2788
1294 ySerIleValAlaGlyGlyGlyAsnThrIleIleAsnleuThrArgAla 1311
2789 GCGCTCCCTATTATCGTTACACCGCACTCGTAGAATCCCGT... 2835
1311 snAspSerLeuThrGlnValAsnSerGlyThrIleThrAlaThrAsnGly 1327
2836 .....TT 2837

1328 TyrSerAlaIleThrThrValAsnGlySerAsnAspProIleThrIle 1344
2838 CAACACGCTGACG.....GMAAACGGC..... 2859
1344 pasThrAlaThrGlyValIleAsnGlyIleAsnProAspAlaProLeu 1361
2860 .....AAATTGACNCTGACAGGACATTCGCTTT 2889
1361 leAsnleuGlyArgGlyTyrAsnPheGlyAsnGlnGlyThr..... 1374
2890 ATGTCGAACTCTTCGCGCTACCGACGACAAATTGACGTGGCGGAAG 2939
1375 .....IleAsnValGlnGlyAspAsnAlaValAlaIleSerGly 1387
2940 TTCGGAAGNACTTACACTTGGCGGTCAACATTAACGACGACGACG 2989
1387 yGlyThrSerSerThrValIleAsnleuValAsnSerGlyThrIleAsnV 1404
2990 TAAAGCTCGATCAA.....TTGACG 3009
1404 alGlyThrGlnGlnGlyGlyAspGlyThrAsnGlyThrGlyLeuIle 1420
3010 GTAGTGAAGGGAAGCAACAAACCGCTTCGGAACCTTAATTTCAC 3059
1421 GlyIleGlySerGlyAsnGlyAsnAlaThrThrIleAsnAsnThrAlaAsp 1437
3060 CCTCAAAACGACACGTCGAT.....GCCGCGCGCTGCGCTTAC 3100
1437 yValIleAsnValTyrAlaAspAspSerThrAlaPheGlyGlyThrL 1454
3101 AACTCATCCGCAAAAGACGCGAGTTCGCGCTG..... 3132
1454 ysaIleIleIleAsnAsnGlyGlyIleAsnleuLeuGlyAspSerGlyCys 1470
3133 .....CATAT..... 3138
1471 AspIleThrAlaProGlyThrThrGlyThrGlnAsnAspHisAsnGlyTh 1487
3138 ..... 3138
1487 rAlaAspIleValIleProAspAlaThrThrAlaProThrGlnGlySer 1504
3139 .....CCGTCAAAGACAGACGCTTTCGACAAA 3168
1504 leProThrProProAlaAspProAsnAlaProGlnGlnLeuSerAsnThr 1520
3169 CTC.....GGCAAGCGAAGCAACCAAAA 3191
1521 IleValGlyThrAsnAlaAspGlySerSerglyThrLeuIleAsnAs 1537
3192 ACAAGCGGAAAGACAGACGCGCAAGCCTTGACGCGCTGATTGCGCGG 3241
1537 nleuValIleGlyAspAsnValGlyValAspThrGlyPheThrSerglyT 1554
3242 GCGCGCATTCGCGCGCAAGACGAAAGCGTTCGCAACCGCGCGCG 3291
1554 hrAlaAspThrThrValAlaValAlaAspAsnAlaPheThrGlySerAsn 1570
3292 GCAGCGCGGGAATAATGTC.....GGCATATGACAGCGGAGAGAAGA 3332
1571 GlnGlyAlaAspAsnIleThrSerThrSerValAlaThrAsnAlaGln 1587
3333 GAAAAAAGCGGTGACGCGGATTAAGACAGCGGCTTGGGCAACAGCGG 3382
1587 ySerGlnAspAlaAspGlyAsnValAspValThrMetThrLysAsnAlaT 1604
3383 AAGCGGAACCCGCGCGGNTACACGCGCTTCGCCGCGCGCGCGCG 3432
1604 yralAspValAlaThrAspSerSerVal..... 1613
3433 CGCGCGGATTTGCCGACACCGACGCGCAACCGCAACTCAACCCACAC 3482
1614 .....SerAspValAlaGlnAla..... 1619
```

```

3483 GCAGCGCAGCTGATNAGCCGTTATCCCAATAGCGTTG..... 3522
      ||| :||: |||: |||: |||
1620 .....LeuAspAlaGlyTyrThrAsnAsnGluLeuYrThrSerL 1633
3523 .....AGTGAATTTCCCGCAGCGCTCAGACGCGTTTTC 3555
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
1633 euAsnValGlyThrThrAlaGluAsnSerAlaLeuLysGlnValSer 1649
3556 GCCGACAGAGAGCATTTGACCGCGTGTCCGACAGCCG..... 3597
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
1650 GtYAlaGln.....AlaThrThrAlaPheArgGlnAlaArgValLeuSe 1664
3597 ..... 3597
1664 rAsnArgPheThrMetLeuAlaAspAlaAlaProGlnIleLysAspGlyL 1681
3598 .....CGCAACCCGCTTTGGACACAGCAGCAGCGGACCAACACATAC 3642
      |||: |||: |||: |||: |||: |||: |||: |||: |||
1681 euAlaPheAsnValAlaAlaLysGlyAspProArgAlaGluLeuGlyAsn 1697
3643 CGTTGGCAA...GATTCGGCGGCTTACCGCAGCAACACGCTGCGC.. 3687
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
1698 AspThrGlnTyrAspMetLeuAlaLeuArgGlnThrLeuAspLeuThrAl 1714
3688 .....CAATCGGTATGCAGAAAAACCTCGGACGCG 3718
1714 aSerGlnAsnLeuThrLeuGluTyrGlyIleAlaIArgLeuAspGlyAspG 1731
3719 GCGCGCTGGCATCTCTTTTCGACACAGCGAGCGAAAAACANTTCGAC 3768
      ||| :||: ||| :||: ||| :||: ||| :||: |||
1731 Ly.....SerLyThrAlaGlyAspAsnGlyLeuThr 1741
3769 GACGGCATCGGCACTCGGACGCGTTCGCCACGCGCGCTTTCGGGCA 3818
      ||| :||: ||| :||: ||| :||: ||| :||: |||
1742 GlyGlyTyrSerGlnPhePheGlyLeuLysSerMetAlaPheAspG1 1758
3819 ATACCGCATCGGC.....AGGTTCGATCATC...GGCATCA 3850
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
1758 u...GlyLeuAlaIArgPheAsnAsnSerLeuArgTyrAspAlaHisAsnLeu 1774
3851 GCACGGCGCGCGGTTTACGACGCGCANTTCNTACAGCGCATCGGAGCG 3900
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
1774 spSerSerArgSerValAlaTyrGlyAspValAsnLysIleAlaAspSer 1790
3901 AAAATCGCGCGCGCTGCTGATTC.....GGCATTCAGGCGCG 3941
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
1791 AspMetArgGlnGlnTyrLeuGluPheArgSerGluAlaLysThrPh 1807
3942 ATACCGCGCGGTTTCGGCGGATTCGCGCATCGAACCGTACATCGGCGAA 3991
      ||| :||: ||| :||: ||| :||: ||| :||: |||
1807 eThrMetMetGly..AspAlaLeuLysValThrProTyrAlaGlyValL 1823
3992 CGCGCTATTTGCTCCAAAAAGCGATTACCCCTACGAAAAACGTAATATC 4041
      ||| :||: ||| :||: ||| :||: ||| :||: |||
1823 ys.....PheArgHisThrMetGluAspGlyTyrLysGluArgSerAla 1837
4042 GCCACCCCGGCTTCGCTTCACCGGNTACG..... 4074
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
1838 GlyAspPheAsnLeuSerMetAsnSerGlyAsnGluThrAlaValAspSe 1854
4075 .....GCGGGCATTTAGCGAGATTTATTCATTAACCGCGCAGACACATNT 4120
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
1854 rIleValGlyLeuLysLeuAspTyrAlaGlyLysAspGlyTrpSerAlaT 1871
4121 CCATCACN.....CCTTATTTTACCCTGTCTCTATACGAGATGCC 4158
      ||| :||: ||| :||: ||| :||: ||| :||: |||
1871 hraIaThrLeuGluGlyGlyPro.....AsnLeuSerTyr..... 1882
4159 GCTTCGGGCAAGTCCGACACGGGTCAATACCGGNTGATGGTCACGGA 4208
      ||| :||: |||: |||: |||: |||: |||: |||: |||
1883 ...SerLysSerGlnArgThrAlaSerLeuGlnGlyAlaAlaGlyGlnSe 1898

```

```

4209 TTTCGGC 4215
      :|||: |||
1898 rPheGly 1900

seq_name: SwissProt_40: YFAL_ECOLI

seq_documentation_block:
ID YFAL_ECOLI STANDARD: PRT: 1250 AA.
AC P45508; P45507; P45506; P39441; P76468; P77487;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yfal precursor.
GN YFAL OR B2233.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=9742617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubaram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiiuchi T.;
RT "K12 genome corresponding to 50.0-66.8 min on the linkage map and
RT analysis of its sequence features."
RL DNA Res. 4:91-113(1997).
RN [3]
RP PRELIMINARY SEQUENCE OF 1-938 FROM N.A.
RX MEDLINE=84272624; PubMed=6087316;
RA Carlson J., Fuchs J.A., Messing J.;
RT "Primary structure of the Escherichia coli ribonucleoside diphosphate
RT reductase operon."
RL Proc. Natl. Acad. Sci. U.S.A. 81:4294-4297(1984).
RN [4]
RP SEQUENCE OF 925-1198 FROM N.A.
RC STRAIN=K12 / EMG2;
RA Estep P., O'Keefe T., Robison K., Church G.M.;
RL Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE OF 1180-1250 FROM N.A.
RC STRAIN=OV6;
RX MEDLINE=88201664; PubMed=2834621;
RA Hussain K., Elliott E.U., Salmond G.P.C.;
RT "The complete sequence of gyraA."
RL Mol. Microbiol. 1:259-273(1987).
RN [6]
RP IDENTIFICATION
RX MEDLINE=96032851; PubMed=7567469;
RA Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C.,
RA Danchin A.;
RT "Detection of new genes in a bacterial genome using Markov models for
RT three gene classes."
RL Nucleic Acids Res. 23:3554-3562(1995).
CC -i- SIMILARITY: TO E.COLI YDEK.
CC -i- CAUTION: REF.3 SEQUENCE DIFFERS EXTENSIVELY FROM THAT SHOWN BY
CC MANY FRAMESHIFTS.

```

```

-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
DR      EMBL: AE000313; AAC75293.1; -
DR      EMBL: D90855; BAA16052.1; ALT_INT.
DR      EMBL: D90854; BAA16050.1; ALT_INT.
DR      EMBL: K02672; -; NOT_ANNOTATED_CDS.
DR      EMBL: U00549; AAA74094.1; -
DR      EMBL: Y00544; -; NOT_ANNOTATED_CDS.
DR      EcoGene: Egl12850; yfal.
KW      Hypothetical protein; Repeat; Signal; Complete proteome.
FT      SIGNAL          1      23
FT      CHAIN           24      1250
FT      DOMAIN          919      948
FT      CONFLICT        28      30
FT      CONFLICT        40      40
FT      CONFLICT        65      66
FT      CONFLICT        431      431
FT      CONFLICT        433      434
FT      CONFLICT        478      478
FT      CONFLICT        773      773
FT      CONFLICT        853      853
FT      CONFLICT        923      924
FT      CONFLICT        948      994
FT      SEQUENCE        1250 AA; 131152 MW; 17F98C05E299FC95 CRC64;
      LRVITG -> LITISC (IN REF. 4).
      PAYOPVIMAKVGGYLNLRANQAFMRERDHDAGGDCQTLN
      POTENTIAL.
      15 X 2 AA TANDEM REPEATS OF [DT]-P.
      AAV -> RGRS (IN REF. 3).
      K -> Q (IN REF. 3).
      LV -> PG (IN REF. 3).
      S -> Q (IN REF. 3).
      AG -> SA (IN REF. 3).
      A -> R (IN REF. 3).
      E -> S (IN REF. 3).
      V -> M (IN REF. 3).
      P -> AT (IN REF. 3).

```

```

alignment_scores:
      Quality: 263.50      Length: 1438
      Ratio: 0.422      Gaps: 63
      Percent Similarity: 43.394      Percent Identity: 19.402

```

## alignment\_block:

US-09-303-518D-651 x YFAL\_ECOLI ..

Align seg 1/1 to: YFAL\_ECOLI from: 1 to: 1250

```

697 AATGGCGTANTTAGTTGAGCGC.....GATGTGCGCA 731
      |||||.....
25 AsnGlyValAlaValAlaThrAspSerGlyThrValAlaValAla 41
      ::::
732 TGGCAACGACTATGGCCCTATGCCGATTCAGAGCGGCGAC.... 777
      ::::
41 AserGlyGlnAlaSerArgGlnSerLeuSerGlyThrGlnAspTrpS 58
      .....
778 .....AGCGGTGCGCAATGTTATTATGACAAACAAACATATAA 819
      ::|||::
58 erlAlaAspGlyGlnTrpLeuValPheSerAspMetThrAsnAla 74
      |||||
820 TGGCTGCTCAGACGAGTTTACAAACCGGCTACCTTATTC..... 861
      |||||
75 SerGlyValAlaValAlaPheLeuGlnGlnGlyAlaGluPheSerLeuLeu 91
      ::|||
862 GCGAGGGAAGGCTTTCACCTGATACGCAAGATTGGTTCTACGATG 910
      ::|||::
91 ocLusnGlnThrGlyMetThrLeuPheAlaAsnThrValThrGly 108
      ::|||
911 ACATTTCAGAGGC.....GAT 927
      ::|||
108 lutyAsnGlnGlyAlaAlaPheAlaAlaGluAsnSerThrLeuAsn 124
      ::|||
928 ACACATACCGTCTNTTGAACCGGCAAGTACGACATTTTTC..... 972
      |||||
125 LeuThrAspValIlePheSerGlyAsnValAlaGlyGlyThrGlyAla 141
      ::|||
973 ....TTTACATCAACAACACGAGTACGGGTACGTAACAGAACCAACG 1018

```

```

      ::|||::
141 alLeuSerSerGlyThrAsnAspThrGlyAlaVal..... 153
      ::|||
1019 AAAAGGINTCCAAATCCAAAGGTTAAAGTACAGACGCGAGCTTTGAC 1068
      |||||
154 .....AspLeuArgValThrAsnAlaMetPheAlaAsn 164
      ::|||
1069 GAATCTTGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 1118
      ::|||
165 AsnIleAlaAsnAspGlyLysGlyAlaAlaIleThr..... 177
      ::|||
1119 TAATCAGTACCGTCCAAAGCTTAACAAACGGTGAACCTTTCTTTATCG 1168
      ::|||
178 .....IleAsnAsnAspValThrLeuSerAspVal 188
      ::|||
1169 ATTACGGCAACGGCAACCTCATCTTATCAACAAACATCAACCAAGCGCG 1218
      ::|||
188 lePheAspAsnGlnAlaIleThrThrSerThrSerThrSerAspGlyAsp 204
      ::|||
1219 GCGCGTTTGTATTTTGAAGTGAT..... 1242
      |||||
205 GlyGlyAlaIleAspValThrAspAsnAsnSerAspSerThrSProse 221
      ::|||
1243 ....TTTACGGTC.....TCGCTGAACCAACGAAACGTGGCAG 1279
      ::|||
221 rGlyThrThrIleValAsnAsnThrAlaPheThrAsnAsnThrAlaGlu 238
      ::|||
1280 GCGCGGCG.....GTT 1290
      |||||
238 lYThrGlyGlyAlaIleThrThrAsnSerValThrAlaProTyThrIle 254
      ::|||
1291 CATATCAGTGAAGACAGTACCGTACTTGAAGTAAACGGCTGGCAAA 1340
      |||||
255 AspIleSerValAspAspSerThrSerGlnAsnGlyValLeuValAs 271
      ::|||
1341 CGACCGCTGTCCAAATTCGCAAGGACGCTGCACGTTCAAGCCAAAG 1390
      ::|||
271 pGluAsnAsnSerAlaAlaGlyThrGlyAspGlyProSerSerAlaAla 288
      ::|||
1391 GGGAA.....AACCAAGCTCGATCAGCGGCGGAC 1422
      |||||
288 lYglyPheMetThrLeuGlyLeuSerGluValThrPheAspIleAlaAsp 304
      ::|||
1423 GGTACAGTCAATTTGATCAGACGAGCAGATTAAGCAAAACAAAGC 1472
      |||||
305 GlyThrLeuValIleGlyAsnThrGluAsnAspGlyAlaValAspSe 321
      ::|||
1473 CTTTAGTGAATCGCTTG.....NTCAGCGCAGGGGTACGGTGCAC 1516
      ::|||
321 rIleAlaGlyThrGlyLeuIleThrIlyThrGlySerGlyAspLeuVal 338
      ::|||
1517 TGAATGCCGATTAAT..... 1530
      |||||
338 euAsnAlaAspAsnAsnAspPheThrGlyGluMetGlnIleGluAsnGly 354
      ::|||
1530 ..... 1530
      ::|||
355 GluValThrLeuGlyArgSerAsnSerLeuMetAsnValGlyAspThrH 371
      ::|||
1531 ....CACTTCAACCCGACAAACTATTTGGCTTTCCGGCGGACGTT 1576
      |||||
371 scYsGlnAspAspProGlnAspCysTrp...GlyLeuThrIleGlySer 387
      ::|||
1577 TGAT.....TTAAGGGCATTCGGCTTCGTTCCACCGTATTCAAAT 1620
      ::|||
387 leAspGlnTyrglnAsnGlnAlaGluLeuAsnValGlySerThrGlnGln 403
      ::|||
1621 ACCGATGAAGGGGATGATTCMCNATCATTAATGCAACAAACATCAAC 1670
      |||||
404 ThrPheValHisAlaLeuThrGlyPheGlnAsnGlyThrLeu....As 418
      ::|||
1671 CGTTACCATTAACAGGAATGAAGTATTACACACCGCATGTGAAGATA 1720
      ::|||

```

418 nileasplaglylgnValThrValAsnGlnGlySerPhealaglyl 435  
 1721 TCAATAGACTTAAATTACAGAAAGAAATGGCTACAGCGTTGGTT... 1767  
 435 leileuglValaglylGlnleuThrIleAlaGlnAsnGlySerIyVal 451  
 1768 .....GGCGAGAA.....GATACGACCAAAACGAA 1793  
 452 leuAlaGlylAlaGlnSerMetAlaLeuThrGlyAspIleValAlaAspAs 468  
 1794 CGGCGCGCTCAACCTTTGTTTACACGCCCGCGAGAGACCGCACCCGNC 1843  
 468 polValAlaVal...LeuSerLeuGlnGlyAspAlaAlaAspLeuThrAla 484  
 1844 TCCTTCCGCGGAGACAAAT.....TTAAACGGCAACATCAGCGAAACA 1887  
 484 euGlnAspAspProGlnSerIleValAlaAsnGlnGlyValLeuAspLeu 500  
 1888 AACGGCAAACTGTTTTTCAGCGGAGACCGACACCGCAACGCTTACATCA 1937  
 501 SerAsp.....PheSerThr 505  
 1938 TTTAGAAGACGGGTGTCAAAAATGGAAGTATC.....CCAC 1975  
 505 ITPrlnSerGlyThrSerIyAsnAspGlyLeuGlyValSerGlySerS 522  
 1976 AAGGAAATCGTGTGGGACAGACGACGTGATCACCAGCTTTAAACGG 2025  
 522 ergIyThrValIleGlySerGlnAsp...ValValAspLeuAlaGlyGly 537  
 2026 GAAATTTCCATATTCAGGCG..... 2046  
 538 AspAsnLeuHisIleGlyIyAspGlyLysAspGlyValIyValValVa 554  
 2047 .....GGCGAGCGGTGATTTCCCGCAATGTTGCCAAAGTGG 2083  
 554 IAspAlaSerAspGlyGlnValSerLeuAlaAsnAsnSerIyLeuG 571  
 2084 AA..... 2085  
 571 IyThrThrGlnIleAlaSerGlyThrLeuMetValSerAspAsnSerIln 587  
 2086 ...GGCGATTCATTTGACCAATCAGCCCAAGATTTTGGTGGC 2132  
 588 leuGlyAspThrHisIyAsnArg.....GlnValIlePheThrAspLy 602  
 2133 ACCGATCAAC.....CATACATCTGTACAGCTTGG 2167  
 602 sGlnGlnSerValMetGluIleThrSerAspValAspThrAspSerA 619  
 2168 ACTGACNGTCTGACAAATTTGTGCAANAANACATTACCGAGATAA 2217  
 619 sPAlaIleGlyHisGlyArgAspIleGlnMetArgAlaAspGlyGlyVal 635  
 2218 GTGATGCTTCATTTGACAGACACNTNAGCGGANTGTNAGCTNNC 2267  
 636 AlalaIAspAlaGlyValAspThrGlnTPrlGlyAlaLeuMet..... 649  
 2268 CATATACGTTNTTNAANCCTCNCNGGCTGCNACACTNANAGCAATC 2317  
 650 .....AlaAspSerSerGlyGln 656  
 2318 TTAGTGAACATGCGATACAGCTTTATACGTACGACACAGCCACCCAA 2367  
 656 IAspGlnAspGlnGlySerThr.....LeuThrIlys 665  
 2368 AACGGCAACCTTAGCTCGTGGGCAATGCCAAGCAACATTTAAATCAAG 2417  
 666 ThrGly.....AlaGlyThrLeuGlnLeuThrAlaSerGlyTh 678  
 2418 CACATTAACGGCAACNCACTGCGTGGGCAATGCTTCATTTAATCTAA 2467  
 678 rThrGlnSerAlaValArgValGlnGlyThrLeuLysGlyAspValA 695

2468 GCAACAAACGCCGCAAAAACGGAGTGTGACGCTTTCGCAACGCTAAG 2517  
 695 IAspIleLeuProIyAlaSerSerLeuThrValGlyAspGlyAlaThr 711  
 2518 .....GCAACGTAAACCATTCGCACTC.....AA 2543  
 712 PheValThrGlyAlaAspGlnAspIleGlnSerIleAspAlaIleSerSe 728  
 2544 CGGCAATGTCTCCCTAGCCGATAGGACAGTATTCATTTGAAACAGCC 2593  
 728 rGlyThrIleAspIleSerAspGlyThrValLeu.....A 740  
 2594 GCTTACCGGCAACTGACGCGGAGAGAGANACAGATTAACCTTAA 2643  
 740 rGlyLeuThrGlyGlnAspThrSerValAlaLeuAsnAla...SerLeuPhe 755  
 2644 GACAGCGAATGACCGCTGCGGTGAGCGAGAA.....TTAGG 2681  
 756 AsnIyAspGlyThrLeuValAsnAlaThrAspGlyValThrLeuThrG 772  
 2682 CAATTTAAC.....CTTGACACAG 2701  
 772 yGlnLeuAsnThrAsnLeuGlnThrAspSerLeuThrIyThrLeuSerAsnV 789  
 2702 CCACCATTAACATCAATTCGCGCTATGCGACAGATGTCGAGCGGCA 2751  
 789 alThrValAsnGlyAsnLeuThrAsnThrSerGlyAlaValSerLeuGln 805  
 2752 ACCGCGAGNCTGCAGACAGCGCGCGCGGTTCGCGCGCTTCCTATT 2801  
 806 AsnIyValAlaGlyAsp..... 811  
 2802 ATCGTTACACCGGCAACTTCGTAAGATCCGTTTACACGCTGACG 2851  
 812 .....ThrLeuThrV 815  
 2852 TAAACGCAATTTGAACNGTCAAGAACATTCGCTTATGTGGAATC 2901  
 815 AlaAsnGlyAspIyThrGlyGlyGlyThrLeuLeuLeuAspSerGlyLeu 831  
 2902 TTCGCG.....TACGAGCGCAAAATTTGACGTGCGGAAAGTTCCGA 2945  
 832 AsnIyAspAspSerValSerAspGlnLeuValMetAsnGlySerThrAl 848  
 2946 AGNACCTTACACCTTGGCGGTCAACAT...ACCGCG...AACGACCG 2989  
 848 agIyAsnThrThrValValAlaValAsnSerIleThrGlyIleGlyLuprot 865  
 2990 TAAAGCTCGATCAATTTGACGCTAGTGAAGGAAAGCAACAAACCGCTG 3039  
 865 hrSerThrGlyIleIyValValAlaAspPheAlaIAspProThrGlnPhe 881  
 3040 TCGGAAAACCTTAAATTTACCCCTGCAAAACGAA...CACGTGATGGCG 3086  
 882 GlnAsnAsnAlaGlnPheSerLeuAlaGlySerGlyTyValAlaMetG 898  
 3087 CGCGTGGCGTTACCACTCATCCGCAAGACGCGGAGATTCGCGCTGATA 3136  
 898 yAlaTyAspIyThrLeuValGlnAspAsnAsnAspIyPtyrLeu... 913  
 3137 ATCCGTCGCAAAAGAACAGAGACTTCCGACAAACTCGGCAAGGAGAGCC 3186  
 913 ..... 913  
 3187 AAAAAACAGCGGAAAAAGAACAGCGCAAGGCTTGACGCGTATGTC 3236  
 913 ..... 913  
 3237 GCGCGGCGGAGTCCCGCAAAAGACAGAAAGCGTTCCGGAACGGGCC 3286  
 914 .....ArgSerGlnGlyValThrProPheSer. 922

3287	GGCNGGAGAGCGGGGAAATATGTCGGATTATATGAGCGGAGGAAAGAA	3336
922	.....	922
3337	AAACGGGTGACGGCGGATTAAGACACAGCGNTTGGCGAAACAGCGGAAC	3386
923	.....ProbaSPRoasPrRoTh.....	929
3387	GGAAACCGGCGGGNTACACAGCGCTTCCCGCGCGCGCNGCGCGCGCC	3436
929	.....	929
3437	GGGATTGGCGGAACGGAGCGCCGACGCGACGACGTCGACCGGACCGAG	3486
930	.....ProaSPRoasPrRoThProaSPRoasPrRoThProaSPRoasPrRoTh	943
3487	CGGACGTGATNAGCGGTTATGCGATATACGCGTTGAGGAAATTTGGCG	3536
944	ProGluPrRoThPrRoAlaValGlnPrRoValLeuAsnAlaValGlyGly	960
3537	CAGCGTCACAGCGCTTTGCGCGTACGAGCGAATTTGGCGCGCGCTTTG	3586
960	YTYuLeuAsnAlaValGAlaVal.....AsnGlnAlaPrHeM	973
3587	CGGAAGCGCGCGCGAAC.....GCGNTTTGGACA	3615
973	eMeGLeuAlaGAGaPrHeNAlaGlyAlaYAsPrGlyAlaThLeuAsnLeu	989
3616	AGCGGATCGCGGAACGCAACAGCTACGCTGCGAAGATTTCCGCGCGCTA	3665
990	ArgValAlaLeuYAlaSprGlyNlSTyTPrAlaAlaGlyGlnLeuAla..	1005
3666	CGCGCAACAAACCGGACGCGCGAATCGGATTCGAGAAACGTCGCGCA	3715
1006	.....GlnHlGAlaSPThSeThValGlnLeuSeGAlaSprLeuPrHeS	1021
3716	CGCGGCGC.....GTGCGATCGTG.....	3735
1021	eYeLAlaTYTrLeuThAsPrGlyAlaTYTrMeLeuGlyAlaLeuAlaYAla	1037
3736	TTTTGCGACACCGGAGCGCAACAGCTGACGAGCGGATTCGCGACATCG	3785
1038	TyLSeLAsPrAsnGlnYAlaSprSeLArgSeLAsMeThPrGlyThPrAlaGAl	1054
3786	GCGACGCGTTGCGGAGCGCGCGCTTTGCGGCAATTCGCAATCGGACGAT	3835
1054	AsPrAsnGlnAsnHlGly.....TyLAlaValGly...	1064
3836	TGCACATCGGACATCACACAGCGCGCGGTTTACGACGCGGACNTCTTACA	3885
1065	.....LeuThSeLSeLTYPrHeGlnHlGAlaYAsnGlnYAla	1076
3886	GACGCG.....ATCGGAGCGCA	3902
1077	GlnGlyAlaTYPrLeuAsPrSeLTYPrLeuGlnTYAlaTYPrHeSeAsPrAs	1093
3903	AATCGCGCGCGCGC.....GTGCGCATTTAC.....GCGATTC	3934
1093	ValSeLSeLglnGlnAlaSprGlyThAsPrAlaTYNlSTyThSeLSeLgylLeI	1110
3935	AGGCAAGATACCGCGCGGTTTC.....GCGGATTCGCG	3969
1110	LeAlaSeLTYPrLeuAlaGlyTYTYGlnTYPrLeuTYGAlaYAlaVal	1126
3970	ATGACACCGCAATCGCGGCAACGCGGATTTGTCGCGAAACGCGGATTA	4019
1127	ThLeuTYPrGlnAlaGlnValAlaTYTYTYGlnGlyAlaGlnGlnAsPrAsPrH	1143
4020	CGCGCTACGAAAC.....GTCAATATCG	4042
1143	eThAlaAlaAsnAlaArgAlaAlaTYAlaSeLglnSeLglnYAlaSprAlaPHeG	1160
4043	CGACCGCGCGTCTTGCGCTTCAACCGGATACCGGCGCGCGCATTTAAGCGAAT	4092

```

1160 Inthr.....ArgLeuGlyLeuHisSerGlu 1168
4093 TATTCATCAACCGCGCACACATNTCCATCNCCTTATTTNAGCCT 4142
1169 Trp.....ArgThrAlaValHis.....ValIleProThrLeuAspIle 1181
4143 GTCC...TTACCGATGCCGCTTCGGGCAAAAGTCCGACACACGGTCAATA 4189
1181 uasnTyrrThiHisAspProHisSerThrGluIleGluGluAspGlySerT 1198
4190 CGCGNCTATTTGGCTCAGATTTCGGCAAAACCGCG...AGTCGGGAATG 4236
1198 hr.....IleSerAspAlaValIleGlyGlnArgGlyCyluIleIysVal 1212
4237 GCGGTAAACGCCGAATC...AAAGTTTCACGCTGCTCNCCTCACGCTGC 4283
1213 GlyValThrGlyAsnIleSerGlnArgValIleLeuArgGlySerValI 1229
4284 CGCGCGCAAAAGNCCGCAACTGGAGCGCACACACAGCGGCGCATCAAT 4333
1229 atpGlnIysGlySerAspAspPheAlaGlnThrAlaGlyPheLeuSerM 1246
4334 TAGGCTACCGCTGG 4347
1246 ecThrValIysTrp 1250
seq_name: SwissProt_40:YEJO_ECOLI
seq_documentation_block:
ID YEJO_ECOLI STANDARD; PRT; 863 AA.
AC P33924; P76450; Q47291;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical outer membrane protein yejo.
GN YEJO OR B2190.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97442617; Pubmed=9278503;
RA Blatter F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Siao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[3]
RP PRESENCE OF AN INSERTION SEQUENCE.
RA Ruod K.E.;
RL unpublished observations (JAN-1994).
CC -1- SOURCE/CDLIR LOCATION: Outer membrane (potential).
CC -1- SIMILARITY: STRONG. TO BORDETELLA PERTUSSIS.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION: THE READING FRAME IS
CC INTERRUPTED BETWEEN CODONS 21 AND 22 BY A 155k INSERTION ELEMENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license@isb-sib.ch.
CC or send an email to license@isb-sib.ch.

```





```

2687 TAAACCTTGCAACGCCACCATTTACACTCAATTCGCCCTATCGCACGAT 2736
      |||  ::  ::  ::|||::|||::|||::|||
419  IeThrThrThrAsnGlnAsnThrLeuLeuAsnLeuAla.....AsnLeu 433
      |||  ::  ::  ::|||::|||::|||::|||
2737 GGTGACAGCGCGCAACCGCGCAGNGTGTGACAGACCGCGCGCGCGCTTC 2786
      |||  ::  ::  ::|||::|||::|||::|||
434  AlameSerAspAlaAsnValIleMetMetAspGluProValThrThrGse 450
      |||  ::  ::  ::|||::|||::|||::|||
2787 GCACCGCTTCCCTATTCGCTTACCGCGCAACTCGTAGAATCCCTTC 2836
      |||  ::  ::  ::|||::|||::|||::|||
450  r.....SerValThr.....AlaSerAlaGluAsn...P 459
      |||  ::  ::  ::|||::|||::|||::|||
2837 TCACACGCTGACGAGTAAACGGCAATTTGACNGTCAAGAACATTCGCC 2886
      |||  ::  ::  ::|||::|||::|||::|||
459  heIleThrThrThrAsn...ThrLeuSerGlyAsnGlyAsnPhetYr 474
      |||  ::  ::  ::|||::|||::|||::|||
2887 TTTATGTGGAACTCTTCGGCTACGACGACAAATGAAAGTGGCGGA 2936
      |||  ::  ::  ::|||::|||::|||::|||
475  MetArgThrAspMetAlaAsnHisGlnSerAspGlnLeuAsnValThrGl 491
      |||  ::  ::  ::|||::|||::|||::|||
2937 AAGTTCCGAAGNACTTACACCTTGGCGGTCAACAATACGGCAACGAAAC 2986
      |||  ::  ::  ::|||::|||::|||::|||
491  yGlnAlaThrGlyAspPheGlyIlePheValThrAspThrGlyAlaSerP 508
      |||  ::  ::  ::|||::|||::|||::|||
2987 CCTTAAGCCTCGATCAATTGACGGGTAGTGGAA.....GGGAAGAACAAC 3030
      |||  ::  ::  ::|||::|||::|||::|||
508  roIaIaIaGlyAspSerLeuThrLeuValThrThrGlyGlyAsp... 523
      |||  ::  ::  ::|||::|||::|||::|||
3031 AAACCGCTGTCCGAACCTTAATTACCTCGCAAAACGAAACGAAAC... 3075
      |||  ::  ::  ::|||::|||::|||::|||
524  .....AlaIaIaPheThrLeuGlyAsnAlaGlyGlyA 534
      |||  ::  ::  ::|||::|||::|||::|||
3076 .GTGATGCCGCGCGCGCTGTACCACTCATCGCAAAAGACGGCGAGT 3124
      |||  ::  ::  ::|||::|||::|||::|||
534  IValaAspIleGlyThrGlyGlyThrLeuLeu..... 545
      |||  ::  ::  ::|||::|||::|||::|||
3125 TCCGCTGTGATATCGGTCAAAAGAACAGACGCTTCCGACAACTCGGC 3174
      |||  ::  ::  ::|||::|||::|||::|||
545  ..... 545
      |||  ::  ::  ::|||::|||::|||::|||
3175 AAGGAGAGAGCCAAAAMAAACAGCGGGAAGAAAGACGCGCAAAAGCTTGA 3224
      |||  ::  ::  ::|||::|||::|||::|||
545  ..... 545
      |||  ::  ::  ::|||::|||::|||::|||
3225 CGCGCTGATTCGCGCGCGCGCGGATGCGCGCGAAGAACAGACGCTTG 3274
      |||  ::  ::  ::|||::|||::|||::|||
546  .....AspAsnGlyAsnHisSerThrSerLeuA 555
      |||  ::  ::  ::|||::|||::|||::|||
3275 CCGAACCGCGCGCGCGCGAGCGGGGAAATGTGGCATTTATGACAGCG 3324
      |||  ::  ::  ::|||::|||::|||::|||
555  IagIu..... 556
      |||  ::  ::  ::|||::|||::|||::|||
3325 GAGGAAGAGAAAAACGGGTGACGGGATTAAGACAGCGCNTTGGCGAA 3374
      |||  ::  ::  ::|||::|||::|||::|||
556  ..... 556
      |||  ::  ::  ::|||::|||::|||::|||
3375 ACAGGCGCAAGGGAAGAACCGCGCGGNTACACGCGCTTCCCGCGCGCC 3424
      |||  ::  ::  ::|||::|||::|||::|||
557  ....AsnAlaGlyAlaGlnIleThrProSerThrThr..... 566
      |||  ::  ::  ::|||::|||::|||::|||
3425 GCNGCGCGCGCGGGATTTGCCGCAACGCGACGCCCAACGCAACCTCAA 3474
      |||  ::  ::  ::|||::|||::|||::|||
566  ..... 566
      |||  ::  ::  ::|||::|||::|||::|||
3475 CCCCACACGCGACCGGACCTGATNAGCGCTTATGCAATAGCGGTTGAG 3524
      |||  ::  ::  ::|||::|||::|||::|||
567  .....AspValLeuAsnMetAlaAlaIaIaGlnProLeu.. 577
      |||  ::  ::  ::|||::|||::|||::|||
3525 TGAATTTCCGCGACGCTCAACAGCGCTTTGCGCGTACAGAGCAAGATTGG 3574
      |||  ::  ::  ::|||::|||::|||::|||
578  .ValPheAspAlaGlnLeuAspThr.....ValArgGluArgLeuG 591
      |||  ::  ::  ::|||::|||::|||::|||
3575 ACCGGGTGTTTGCCGAAGACCGCGCGCAACGCGNTTGTGCAAGACGATC 3624
      |||  ::  ::  ::|||::|||::|||::|||

```

---

```

591  IySerValIySelYalYalSerTyAspThrAlaMetThrPseSerAlaIle 607
      |||  ::  ::  ::|||::|||::|||::|||
3625 CGGNACACCAACACACTACCGCTTCGACAGATTTCCGCGCTTACCGCAACA 3674
      |||  ::  ::  ::|||::|||::|||::|||
608  AsnThrArgAsnAsnValThrThrAspAlaGlyAlaGlyPheGluGlnTh 624
      |||  ::  ::  ::|||::|||::|||::|||
3675 AACCGACCTGGCGCAATCGGTATGCGAAGAAACCTC.....G 3712
      |||  ::  ::  ::|||::|||::|||::|||
624  rLeuThrGlyLeuThrLeuGlyIleAspSerArgPheSerArgGluGlus 641
      |||  ::  ::  ::|||::|||::|||::|||
3713 GCAGCGCGCGCGCGCATCCCTGTTTTCGCAACACCGGACGCAAGAACANC 3762
      |||  ::  ::  ::|||::|||::|||::|||
641  eSerThrIleArgGlyLeuIlePheGlyTySerHisSerAspIleGly 657
      |||  ::  ::  ::|||::|||::|||::|||
3763 TTCGACGACGCGCATCGGCACTCGGACAGGCTTCCACAGCGCGCTTTT 3812
      |||  ::  ::  ::|||::|||::|||::|||
658  PheAspArgGly..... 661
      |||  ::  ::  ::|||::|||::|||::|||
3813 CGGCAATACGCAATCGGACAGTTCGACATCGCATCAGACGCGGCGCGG 3862
      |||  ::  ::  ::|||::|||::|||::|||
662  .GlyIySelYalYalAspSerTyThrLeuGlyAlaTyAlaIaGlyTrpG 678
      |||  ::  ::  ::|||::|||::|||::|||
3863 GTTTTACGCGCGCGCATCTNTACAGCGGC..... 3891
      |||  ::  ::  ::|||::|||::|||::|||
678  IuhIsgIuhAsnGlyAlaTyValAspGlyValAlaYalYalAspArgPhe 694
      |||  ::  ::  ::|||::|||::|||::|||
3892 .....ATCGGAGCGCAAAATCCGCGCGCGCTGCTGATTAAGC.. 3930
      |||  ::  ::  ::|||::|||::|||::|||
695  AlaAsnThrIleHisGlyLysMetSerAsnGlyAlaThrAlaPheGlyAs 711
      |||  ::  ::  ::|||::|||::|||::|||
3931 .....ATTCAGCGCAGCATACCGCGCGG 3952
      |||  ::  ::  ::|||::|||::|||::|||
711  pTyAsnSerAsnGlyAlaGlyAlaHisValGluSerGlyPheArgTrpY 728
      |||  ::  ::  ::|||::|||::|||::|||
3953 GTTTCGCGGATTCGCGCATCGAACCGGTACATCGCGGCAACGCGCTATTTC 4002
      |||  ::  ::  ::|||::|||::|||::|||
728  alaSpGlyLeuThrPseValArgProTyLeuAlaPheThrGlyPheThr 744
      |||  ::  ::  ::|||::|||::|||::|||
4003 GTCCAAAAGCGGATTCACCGCTACGGAAGGTCAATTCGCCACCCCGG 4052
      |||  ::  ::  ::|||::|||::|||::|||
745  ThrAspGlyGlnAspTyThrThrLeuSerAsnGlyMetArgAlaAspValGl 761
      |||  ::  ::  ::|||::|||::|||::|||
4053 T...CTTGCCTCAACCGNTACCGMGCGGCGCATTAAGGAGATTTATCAT 4099
      |||  ::  ::  ::|||::|||::|||::|||
761  YAsnThrArgIleLeuArgAlaGlnAlaGlyThrAlaValSerTyHisM 778
      |||  ::  ::  ::|||::|||::|||::|||
4100 TCAAACCGCGCGCAACACATTCATCAGNCCTTATTTAGCCTGTCTAT 4149
      |||  ::  ::  ::|||::|||::|||::|||
778  etAspLeuGlnAsnGlyThrThrLeuGlnIuProThrPleuYalAlaIaVal 794
      |||  ::  ::  ::|||::|||::|||::|||
4150 ACCGATGCGCGCTTCGGGCAAGTCCGAAACGCGGATCAATCCGCGCNTATT 4199
      |||  ::  ::  ::|||::|||::|||::|||
795  ArgGlnIuTyAlaIaAspSerAsnGlnValYalYalAsn..... 807
      |||  ::  ::  ::|||::|||::|||::|||
4200 GGTCAAGATTTTCGCGCAAAACCGCAAGTGGGATGGGCGGTAAAC.... 4245
      |||  ::  ::  ::|||::|||::|||::|||
808  ....AspAspGlyLysPheAsnAsnAspValAlaGlyThrSerGlyY 822
      |||  ::  ::  ::|||::|||::|||::|||
4246 .....GCCGAATCAAAAGTTTC.....ACGCTGTCNCTCAC 4278
      |||  ::  ::  ::|||::|||::|||::|||
822  aTyArgIuhAlaGlyIleArgSerPheThrProThrLeuSerGlyHis 838
      |||  ::  ::  ::|||::|||::|||::|||
4279 GTTCGCGCGCGCAAGNCCGCAACTGGAAGGCAACACGCGCGGCGCAT 4328
      |||  ::  ::  ::|||::|||::|||::|||
839  LeuSerValSerTyGly.....AsnGlyAlaGlyA 849
      |||  ::  ::  ::|||::|||::|||::|||
4329 CAAA 4332
      |||  ::  ::  ::|||::|||::|||::|||
849  IGIu 850
      |||  ::  ::  ::|||::|||::|||::|||

```

seq\_name: SwissProt\_40:YDBA\_ECOLI

```

seq.documentation_block:
ID YDBA_ECOLI STANDARD; PRT; 2003 AA.
AC P33666; P76087; P76088; P76856; P76857; P76859;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ydbA.
GN YDBA OR B1401/B1405.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=9742617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alpa H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 464-2003 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92190338; PubMed=1665988;
RA Moszer I., Glaser P., Danchin A.;
RT "Multiple IS insertion sequences near the replication terminus in
Escherichia coli K-12.";
RL Biochimie 73:1361-1374(1991).
CC -1 SIMILARITY: TO S.TYPHIMURIUM ORF NEAR CYS6 (AC P25928).
CC -1 CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR
THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT
BETWEEN AMINO ACIDS 839 AND 840.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000237; AAC74483.1; ALT_SEQ.
DR EMBL; AE000237; AAC74487.1; ALT_SEQ.
DR EMBL; D90778; BAI15009.1; ALT_SEQ.
DR EMBL; D90778; BAI18880.1; ALT_SEQ.
DR EMBL; D90779; BAI18881.1; ALT_SEQ.
DR EMBL; X62680; -; NOT_ANNOTATED_CDS.
DR Ecogene; EG1307; ydbA.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 489 489 I -> V (IN REF. 2).
FT CONFLICT 495 495 I -> V (IN REF. 2).
SQ SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;

```

```

alignment_scores:
Quality: 242.00 Length: 1528
Ratio: 0.358 Gaps: 77
Percent Similarity: 44.241 Percent Identity: 19.241

```

```

alignment_block:
US-09-303-518D-651 x YDBA_ECOLI ..
Align seg 1/1 to: YDBA_ECOLI from: 1 to: 2003
82 TTAGCCATATGCGTGTGGCGATTTCCCGAAGCTTGGCGGAGCA 131
7 LeuSerAlaCysIleAlaLeuAlaLeuSerGlyGlnGlyTrpAlaIle... 22
132 CACTTATTTCGGCATCAACTACCAATCTATCCGACTTGGCGAATAA 181
23 .....AspIleThrGlu.... 26
182 AAGCGAAGTTTGCATGCGGCGGAA.....GATATTGAGCTTATAC 222
27 ..ValIGluThrThrGlyGlyIleValSerAsnThrValThrCysPro 42
223 AACAAAAAGGAGGAGTTGTCGGCAATCAATGACAAAGCCCGATGAT 272
43 AlaSprProGlyLysLeuSerProGluGluLeuLysArgLeuProSerGI 59
273 TGATTTTCTGTGCTGTCGCGCTAAC.....GGCG 301
59 uCysSerProLeuValGluGlnAsnLeuMetProTrpLeuSerThrGlyA 76
302 TGGCGGATGTTGGGCGGATCAATATATGTGAGCGGTGCGACATTAACGGC 351
76 IAlaIAlaLeuIleThrAlaLeuAlaValAlaGluLeuAsnAspAsp 92
352 GCGTAT.....AACACGTTGATTT..... 372
93 AspHisHisHisAspAsnAsnSerProLeuProThrProProAspAs 109
373 .....GTCGCGAAGCAAGNA 388
109 pGluSerAspAspThrProValProProThrProGlyAspValIleI 126
389 ATCCCGATGACACCGCTTTTCTTACCAAAATTGGAAAGCAAAVATAT 438
126 IeProAspAsp..... 129
439 AAGCGTGAATTCACAC.....CCTTACACGCGCA 470
130 ...ProAspAspThrProThrProProLysProValSerPheAsnAs 145
471 TTACCATATGCGCGCTTGCATTAATTGTGCACGATGACGACCTCTCG 520
145 pValIleLeuAspLysThrGluLysThrLeuThrIleArgAspSerValp 162
521 AAATGACGAGTACATGAGGGGGAATACCTATCCGATAAAGAAATAT 570
162 herThrTyThrGluAsnAlaAspGlyThrIleSerLeuGlnAspSerAsn 178
571 CCCGAGGTGCCGATCGCTCAGACACCATTCATTGGCTATGATGA 620
179 GlyArgLysAlaThrIle.....AsnLeuTrpGlnIle...As 190
621 TGACAAACGCGGATTTATCTTACTCCGCGCATGCTTAATTCGCGCA 670
190 pGluAlaAsnAsnThrValAlaLeuGlnGlyVal...SerAlaAspLys 206
671 ATACACATATGCGGTTGGGGAATATAGCCGTANTTACTTTGAGCGGC 720
206 IatThrLysTrpGln...TyrAsnHisAsnGlyGluLeuValIleThrGly 221
721 GATGTGCGCATGTCACAGCATATGCGCTATGCGGATGAGGTGCGGC 770
222 AspAsnAlaThrValAsnAsnAsnGlyLysThrThrValAspGlyLys.. 237
771 AGCGACACGCGTTCGCATGTTTATTATGACAAACAAACATTAAT 820
238 .....AspSerThrGlyThrGlu.. 243

```

```
821 GGCTGCTCAACGGA.....GTTTACAAACCGGC...TACCT 855
      ::::|
244 .....lleasnnglyasnnglyvallelelnasplasyaspleuas 258
      ::::|
856 TATTCGGCGAGGAAACGGTTCCAGCTGATACGCAAGATTGGTTCA 905
      ::::|
259 valsercilyglyhnsnglyleaspile..... 268
      ::::|
906 CGATGACATTTCAGAGCGATACACATCCGCTCTTTTGAACCGCGCA 955
      ::::|
269 .....Thrglyaspseralathrval..... 275
      ::::|
956 GTAACGACATTTTTCCTTACATCCAAACACACGCTACGGGTACGTA 1005
      ::::|
276 .....Aspsnnglythrmethrval 283
      ::::|
1006 ACAGAAACACGAAAGGNTCCATCCAAAGCTTAAGTACAGACAGT 1055
      ::::|
284 ThrAspProglu.....Sermetcglylelelnleas 294
      ::::|
1056 CGGACTGTTGACGAATCTTGAATGAATGATAAGAACAGCTTAGC 1105
      ::::|
294 pgl.....Asplyalalevalasnsnglygluserthrllet 309
      ::::|
1106 CGGACAGGGGTGTTAATCAGTAC.....CGTCCAAAGTTA 1140
      ::::|
309 hrasnnglythrnglythrnglyleasnnglyaspsalathrvalaasn 325
      ::::|
1141 AACACGGTGAACCT.....TCCTTTAT 1166
      ::::|
326 Asnasnnglythrthrvalasplylyaspsertthrglythrval 342
      ::::|
1167 CGATTACGCAACGCAACATCTTATCAACAAACATCAACCAAGCGC 1216
      ::::|
342 eaasnnglyasnnglyvallelelnasplasyaspleuaspsalserc 359
      ::::|
1217 CGGCG.....GGTTGATTTTGAAGTATTTACGGTCCGCTGAA 1260
      ::::|
359 lyglylyhnsnglyleaspilethrnglyasp...seralathrvalasp 374
      ::::|
1261 AACACGAACAGTGG.....CAAGCGCGGGGCTCATAT 1295
      ::::|
375 AsnlysglythrmethrvalthrAspProgluserlleglyllelna 391
      ::::|
1296 CAGTGAAGACAGTACGCTTACTTGAAAGTAAACGCGTGGCAACGACC 1345
      ::::|
391 lAspglyaspglalaval.....Valasnsnglyglusera 405
      ::::|
1346 GCCTGTCGAATCGCAAGACGCTGACGCTTCAAGCCAAAGGGCA 1395
      ::::|
405 lalthrAsnnglythrnglythr.....Glnleasnnglyasp 418
      ::::|
1396 AACCAAGGCTCGATCAGCGTGGCGACGCTACAGTCATTTTGGATCAGA 1445
      ::::|
419 AsplalthrAlaasnnsnglylythrthrval..... 430
      ::::|
1446 GGCAGACGATTAAGCAAAACACCTTAGTGAATCGCTGNTCA 1495
      ::::|
431 .....AspglylyaspsertthrglythrvallealaglyAsna 444
      ::::|
1496 GCGGCGAGGGGTACGTGCACTGAATCCGATTAATCAGTTCAACCCGAC 1545
      ::::|
444 snnglylys..... 446
      ::::|
1546 AAACCTATTTCGGGCGGCGAGCTTTGATTTAAAGGCGATTC 1595
      ::::|
447 .....Vallelnasplasyaspleuaspsalsercglygl 459
      ::::|
1596 GCTTTCGTTCCACCGTATCAAAATACGATGACGGGCGATGATGNCN 1645
      ::::|
459 y.....Hlsnglyleaspilethrnglyaspseralathrval..... 471
      ::::|

1646 ATCATATGCCACAACACATCCACCGTTTACATTAAGAGGATGAAGT 1695
      ::::|
472 .....AspsnnglythrmethrvalthrAspProgluser 484
      ::::|
1696 ATT...ACACAACCGAGTGTAAAGATATCAATAGACTTAATTACACAA 1742
      ::::|
485 lleglyllelnleaspglyaspglalalevalasnsnglygl 501
      ::::|
1743 AGAATTCCTTACAAACGGTGTGGCGAGAAA.....GATA 1780
      ::::|
501 userthrlethrAsnnglythrnglythrnglyleasnnglyasnspa 518
      ::::|
1781 CGACCAAAACGAAACGGCGCTCAACCTTTTACACGCGCGCAGAA 1830
      ::::|
518 lathrAlaasnnsnsercilylysthrthrval.....Aspglylys 531
      ::::|
1831 GACCGACCCCGCTGCTTCCGCGGA.....ACAAATTTAAA 1868
      ::::|
532 AspserthrnglythrlylealaglyAsnlelelylevalasleuas 548
      ::::|
1869 CGGCAATACGCAACAAACGCGCAAACTGTTTTCAGGCGGACAGCA 1918
      ::::|
548 pglserleuthrvalthrngly..... 556
      ::::|
1919 CACCGCACGCTTACATCATTTAGAAACGGGTGTCAAAATGGAAGT 1968
      ::::|
557 ..Alahsnglyvalgluasnlelelyasp..... 565
      ::::|
1969 ATCCCAAGAGAAATGCTGTGGACACACGATCAGTCAACCGCACGT 2018
      ::::|
566 .....Asnglythrvalasnsnglyaspillevalvalseraspth 580
      ::::|
2019 TAAACGGGAATTTCCATATTTACAGGCGGCGAGCGGTGTTTCCCGCA 2068
      ::::|
580 rglserlelelyalvalleasnnglygluAlathrvalsernt 597
      ::::|
2069 ATGTTCCAAAGTGAAGCGGATTCATTTGACCAATCACGCCCAAGCA 2118
      ::::|
597 hr.....GlyaspvalAsnvalserasnnglyalathrgly 608
      ::::|
2119 GTTTTGGTGTGCGACGCAAT..... 2139
      ::::|
609 ...PheserlethrthrAsnsercilylyvalserleualaglyser 624
      ::::|
2140 ..CAAGCCATCAATCTGTACACGCTTGACGTGACNGCTGTGCAAT 2188
      ::::|
624 tglvalglyaspPheserthrnglyvalaspleuasnglyAsnasnns 641
      ::::|
2189 GTGTGGAANAANCATTACGACGATTAAGTATGCT...TCATTGACT 2235
      ::::|
641 erValthrleualalalylaspleuasvalvalaglylnlysalathr 657
      ::::|
2236 AAGACNACNTNAGCGCANTGTAGNCTNNCCATNACNTNNTTAAA 2285
      ::::|
658 Glylleasnvalserglyasp..... 664
      ::::|
2286 NCTCNCGGCGNTGNNACCTNMAAGCAATCTTAGTCAATAGCGGATA 2335
      ::::|
665 ...Alasnthrvalasnlethrnglyasnvallevalasplyaspl 680
      ::::|
2336 CACGT..... 2340
      ::::|
680 ysthrAlaaspsalalaglyutrrpPhesaproserValglyle 696
      ::::|
2341 .....TATACATGACCCACACGCGCAACCAAGCGCAACTTACCT 2384
      ::::|
697 AsnValtyrglySeraspasnvalThrleuasplylysthrthrva 713
      ::::|
2385 CGTGGCAATGCCCAACCAATTAATCAAGCACATTA...AACGCA 2431
      ::::|
713 lvalseraspsergluvalthrserarglnserAsnleuPhesaplys 730
      ::::|
2432 ACNCA...TCGNTTCGGCAATGCTTCATTTAATCAACCAACAGCC 2478
      ::::|
```





```
1112 1GlnGlyValValLysAlaAsnThrLeuSerThrAspSerLys 1129
1283 CGGGCGTTCAATATCATGTAAGACAGTACCGTTACTTGGAAAGTAACGCC 1332
1129 erAlaValThrPheThrAspProValValThrGlyAlaLeuAspAsn 1145
1333 GTGGCAACGACCGCTGTCCAAATCGGCAAGACGCTGCACGTTCA 1382
1146 ThrGlyAsn 1148
1383 AGCCAAAGGGAACCAAGCGTCGATCAGC...GTGGCGACGCTACAG 1429
1149 .....AlaAsnAsnGlyLeuValThrPheThrGlyAsnSerThrV 1162
1430 TCATTTTGATCAGCAGCAGCATTAAGCAAAAAACAAGCTTTAGT 1479
1162 a.....ThrGlyAspIleGlyAsnThrAsnAlaLeuValA 1173
1480 GAAATCGCTGNTCAGCGCGGAGGTACGTCGACATGCAATGCCGATA 1529
1174 ThrValAsnValGlyAlaGly.....IleThrLeuGlnAlaGlyG 1187
1530 TCAGTTCAACCCGCAAACTCTATTTCGCTTCGC..... 1566
1187 YSerLeuAlaAlaAsnAsnIleAspPheGlyAlaArgSerThrLeuG 1204
1567 ..GGCGGACGTTGGATTAAACGGCATTCGCTTCGCTCCACCGATT 1614
1204 heAsnGlyProLeuAspGlyGlyAlaIleProGlyTyr..... 1218
1615 CAAATACCGATGAAGGCGATGATGNCNATCATATATGCC..... 1656
1219 .....PheLysGlyAlaIleAlaAsnGlyAsnAsnAlaIleLeu 1232
1657 .....ACACACATCCACCGTTACATT..... 1680
1232 nValAsnThrLysLeuLeuThrAlaSerHisLeuThrIleGlyThr 1249
1681 .....ACAGGAATGAAGTATTACACACGAGGAGGT 1713
1249 IagIuIleAsnIleGlyAlaGlyAsnLeuPheThrIleAspAlaSer 1265
1714 AAGATATCAATAGACTTAATTACAGCAAAAGATTGCTACACGGTTG 1763
1266 GlyAspValThrIleLeuAsn.....AlaGlnAsnIleAs 1277
1764 GTTGGCGAGAAAGATACGACCAAAAGACGCGCGCTCAACCTTGT 1812
1277 nPheArgAlaArgAspSerValLeuValLeuSerAsnLeuThrGly 1294
1813 .....TACCAACCCGCGCGCAAA 1830
1294 IyValAsnAsnIleLeuLeuAlaAlaAspLeuValAlaProGlyAla 1310
1831 GACCGCACCCGCTGCTTTCGCGGAGCA.....AATTAAACGG 1871
1311 GlnGlyThrValValPheAsnGlyGlyValAsnGlyLeuAsnValG 1327
1872 CAACATCAGCAACA..... 1887
1327 rAsnValAlaGlyThrAlaArgAsnIleGlyAspGlyGlyAsnLys 1344
1888 ..AACGGCAACTGTTTTCAGCGGACGACGACGACGCGCTACAT 1935
1344 heAsnThrLeuLeuIleTyAsnAlaValThrIleThrAspAspVal 1360
1936 CATTAGAACCGGTGCTCAAAATGAGAGTATCCACAGAGAGAAAT 1985
1361 .....LeuGlnGlyIleGlnAsnValIle 1369
1986 CGTGGGACACGACTGATCMAACGACG...TTTAAAGCGGAATAAT 2032
1369 eAsnLysAsnAlaAspPheThrSerSerThrAlaPheAsnAlaGlyAla 1386
2033 TCCATATTACAGCGCGGACGCGGTATTTCCCGCAATGTCCCAAGG 2082
1386 IeGlnIleAsnAspAlaThrTyThrIleAspAlaAsnAsnGlyAsn 1402
2083 GAA.....GGCATGNCATTTGACCAATCAAGCCACGCA..... 2118
1403 AsnIleProAlaGlyAsnIleGlnPheAla...HisAlaAspAlaG 1418
2118 ..... 2118
1418 uValLeuGlnAsnSerSerGlyAsnAspArgThrIleThrLeuGlyAla 1435
2118 ..... 2118
1435 snIleAspProAspAsnAspAspGlyIleValIleLeuAsnSerVal 1451
2119 .....GTTTGGTGTGCG 2132
1452 ThrAlaGlyLysLysLeuThrIleAlaGlyGlyLysThrPheGlyG 1468
2133 ACCGATCAAAAGCCATACATC.....GTACAC 2161
1468 a...HisLysLeuGlnThrIleLeuPheLysGlyAlaGlyAspCys 1484
2162 GTTCGACCTGGACN.....GCTCGACAAAT 2187
1484 hrAlaGlyThrThrPheAsnThrThrAsnIleValLeuAspIleThr 1500
2188 TGTGTGGAANAANCATTACCGACGATAAAGTGAT..... 2223
1501 GlnLeuGlnLeuGlyAlaThrThrAlaAsnValValLeuPheAsnAs 1517
2224 ..GCTTATTGACTAAGACGACGACNTNAGCGC..... 2253
1517 aValGlnLeuThrGlnThrGlyAsnIleGlyGlyPheLeuAspPheAs 1534
2253 ..... 2253
1534 IalysAsnGlyMetValThrLeuAsnAsnAsnValAlaGlyAla 1550
2254 .....ANTGNAGCTNNCCNA 2270
1551 ValGlnAsnThrGlyGlyThrAsnAsnGlyThrLeuIleValLeuG 1567
2271 TNACGNTNNTTNAANCCTCMCGCGCTGCGNCCACTN.....AANG 2311
1567 aSerAsnLeuAsnArgValAsnGlyIleAlaMetLeuLysValGlyAla 1584
2312 GCATCTTAGTGCAAT.....GGCGATACAGCT 2340
1584 LysAsnValThrIleAlaLysGlyGlyLysValLysIleGlyGlnI 1600
2341 TATACAGTCAAGCCCAACGCCCAAAAGGCAACCTTAGCTGCTGG 2390
1601 GlyThrGlyThrAsnThrLeuThrLeuProAlaHisPheAsnLeuThr 1617
2391 CAATGCCCAAGCAACATTTAATCAAGCC...ACATTAAAC..... 2427
1617 YSerIleAsnLysThrGlyGlyGlnAlaLeuLysLeuAsnPheMetAs 1634
2428 .....GGCACC 2433
1634 IyGlySerValSerGlyValValGlyThrAlaAlaAsnSerValGly 1650
2434 NCATCGNTTCGGGCAATGCTTCATTTAATCTAAGCAACAGACGCC..... 2478
1651 IleThrThrAlaGlyAlaThrSerPheAlaSerSerValAsnAlaLys 1667
2479 .....GCACAAA 2485
1667 yThrAlaThrLeuGlyGlyThrThrSerPheAlaAsnThrPheThrAs 1684
```

2486 ACGGCACTGTGACGCTTTCCGACACGCTTAAGCAAACTAAGCATTC 2535  
:::|||||  
1684 hrclyalvalthrleu.....Alalysglyserilethrserphe 1697  
:::|||||  
2536 GCACCTCAAGGCAATGTCTCC.....CTAGCGAATTAAGCAAGTATTC 2579  
||| ||| ::::|  
1698 Alalysasnvalthralthrserphevalalasnseralathrille 1714  
:::|||||  
2580 TTTTGAACAACAGC...CGCTTACCGGACACTCAAGCGGACAGCAAGANA 2626  
:::|||||  
1714 npheserasnserleualalpheasnserasnlierhrgly..... 1727  
2627 CAGCATTAACCTTAAAGACAGCAAGTGAACCTGCCCTCAGCAGCGGAA 2676  
1728 .....glylythrthr 1731  
2677 TTAGCAATTTAAACCTTACAAACGCCACCATTAACATTCGCGCTA 2726  
||| ::|||::: ::::|  
1732 leu...Thrleuglyalasnvalthrtythr..... 1742  
2727 TCGCCACGATGCTGACAGCGGCAAAACCGGACAGNGTGCAGACGCGC 2776  
||| ||| ::|||  
1743 .....glythrlyserpetherasphr... 1750  
2777 GCCGCCGCTTCGCGCGCTTCCATTATTCGTTACACCGCAACTTCGCTA 2826  
||| ::|||  
1751 .....LeuthrleuasnthrthrPheaspglyalala 1761  
2827 GAATCCCGTTTCAACACGCTGACGGAAGCAAA...TTGAACNGTCA 2873  
:::|||||  
1762 lysserglyalasnleuleuileysersglyserthrleuaspleuse 1778  
2874 AGGAACATTCCTCGCTTATGTCCGAACCTTCGCTACCGACGCACAAT 2923  
:::|||||  
1778 rglvalserthrleualaleuvalthralthrasnphesapmeta 1795  
2924 TGAAGCTGGCGAAGATTCCGAAGNACTTACACCTTGCGGCTCACAAT 2973  
:::|||||  
1795 snasnlierproasphrlythrlythrValilleseralaglunhrala 1811  
2974 ACCGGCAAGAACCGGTAGCCTGCATCATTAAGGTAGTGAAGGAA 3023  
||| ::|||  
1812 Glyglyleuuleysprothrserlysluasnvalyslierthleasnas 1828  
3024 AGACAACAACCGCTGTCC.....GAACAACCTTAATTACCTTCG 3064  
:::|||||  
1828 naspsnasrghphevalasphethrPheaspalaserthrleuthrleup 1845  
3065 AAAACGACACGCTCGATGCGCGGCGGTGACCTTACCAACTCATCCGCAAA 3114  
||| ::|||  
1845 healaglulaspillealalaaspval.....lle 1854  
3115 GACGGGAGATTCGCGCTGCATTAACCGGCAAGAACAAGACTTCGCA 3164  
||| ::|||  
1855 aspglyasphealaproglycylproleu..... 1864  
3165 CAAACTCGCAAGCAAGCAAGCAAAACAGCGGAGAAAAAGCAACGCGC 3214  
||| ::|||  
1865 .....Alasnlierproasnala 1871  
3215 AAAGCTTGACCGCTGATTCGCGGCGCGGATGCGCGAANAAGCA 3264  
:::|||||  
1871 laasnlierlyserleugluuemetgluaspalaproasnlyser 1887  
3265 GAAACGTTGCCGAACGCGCGCGGCGGAGGGAATGTGCGCAT 3314  
:::|||||  
1888 Asp.....Alaarglinala...Pheasnasnphleglyle 1898  
3315 TATGCAGGCGGAGAGAAAAACGCGGTGCGAGATTAAGACACG 3364  
:::|||||  
1898 uethrthrProleuGinglu.....Alaspalathrthr 1910

3365 CNTTGGCGAAACAGCGGAAAGCGGAACCGCGGTTACACCGCCTTC 3414  
|||  
1910 lsleulle..... 1912  
3415 CCGCGCGCGCGCNGCGCGCGGATTTGCCGCAACCGCGCCCAAC 3464  
:::|||||  
1913 .....Glnaspvallalysproser..... 1919  
3465 GCAACTCAACCCCAACCGGACGCGACTGATNAGCGTTATGCCATA 3514  
1920 .....Aspthrillealalavalasnasng 1928  
3515 GCGGTTGAGTGAATTTCCGCAAGCTCAACAGCGTTTTCGCGTACAG 3564  
:::|||||  
1928 Invalvalala.....Serasnlieraserasnlierhralaleuasn 1942  
3565 GACCAATTTGACCGCGCTTTCG..... 3588  
1943 AlaargmetasplysvalGlnserglyasnlysglyprovalsersergl 1959  
3589 .....GAAGACCGCGCAACGCGGTTTGACACAGC...NGCATCC 3625  
1959 yaspgluaspmetaspalalysphleglyalatrplleserProphervalg 1976  
3626 GGNACACCAAAACACTACCGTTCCGCAAGATTTCGCGCTTACCGCAACAA 3675  
:::|||||  
1976 lysasnlatrhnglnlysmecysasnserlierseerglytylserasph 1992  
3676 ACCGACCTGCGCCCAATCGT.....ATGCAGAAAAACCTCGG 3713  
||| ::|||  
1993 Thrthrlyglythrllieglypheaspglyphealaserasplenal 2009  
3714 CACGCGG.....CGCTGCGGACATCGTGTTCGCACA 3745  
2009 aleuglyleualatyrthrargalaaasphrasipllelyleuileysna 2026  
3746 ACCGACCGAAGAACNCTTGCAGACGCGCATCGGCAACTCGGACGCGCTT 3795  
||| ::|||  
2026 snlysthr.....glyasplyasnlysluasnglyluser 2036  
3796 GCCCAGCGCGCGCTTTCGCGCAATACGCGATGCGGACGATTCACATCGG 3845  
2037 Asnlierthrleuylrglyleuylrlyasnvalprotyrlysluasneuph 2053  
3846 CACGACGACGCGCGGCTTACAGCGCGCATNTTCAGACGCGATCG 3895  
2053 evalglualallealaserlyserasp..... 2062  
3896 GAGGCAAAATCCG.....CGCGCGCTGCATTAAGCATTCAG 3936  
2063 ..AsnlyslleargserlyserArgarvalillealathrthrleuglu 2078  
3937 ..GCACGATACCGCGCGCTTTCGCGGATGTCGATGCAACCGCATCAT 3983  
2079 Thrvallelytyrghlnthrilaasnnglyllystlylserlsuertyrth 2095  
3984 CCGGCAACGCGCTATTTCGTCCAAAAAGCGGATTAACCGGTAC..... 4026  
2095 rgllyglnleu.....Metalaaglytyrthrlymetmetcp 2107  
4027 ..GAANAAGTCAATATTCGCCACCCCGGCTTCGCTTCAACGNTACGN 4074  
2107 rogluasnlleasnleu...ThrPro...leualaglyleuarglytyr... 2120  
4075 GCGGCAATTAAGCAGATTAATTCATTCAAACCGGCGCACACATNTTCAT 4124  
2121 Serthrilleys...Asplysglytyrlysgluthrly..... 2132  
4125 CACNCTTATTNAGCTG.....TCCATA 4150  
2133 ..ThrthrtyrghlnasnleuthrvallysglyllyasnlyrlyasnthrPhea 2149  
4151 CCGATGCCGCTTCGGCGAAGTCCGAACACGCGCTCAATAC..... 4191



```

2149 spg1yleuLeuclYalalysValserSerAsnIleAsnValasnglu1le 2165
4192 .....GCGTATGGCTCGATTCGATTCGCGCAAC 4220
2166 VALLEUTHrProgluLeuTrAlaMetValAspYrAlaPheLysAsnLy 2182
4221 CCGCAGTCGCGAATGGCGCGTAACCGCGAATCAAGTTTCACG 4266
2182 sValserAla.....IleAspAlaArgLeuGlnGlnIleThr 2194.

seq_name: SwissProt_40:YDEK_ECOLI

seq_documentation_block:
ID YDEK_ECOLI STANDARD; PRT; 1325 AA.
AC P32051; P76140; P77168;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipoprotein ydek precursor (ORF).
GN YDEK OR ORF OR B1510.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
CX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isoo K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takada J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 595-1325 FROM N.A.
RX MEDLINE=94100243; PubMed=8274505;
RA Cartwright P.J., Timms M.W., Lithgow T., Hoef P.B., Hoogenraad N.J.;
RT "An Escherichia coli gene showing a potential ancestral relationship
to the genes for the mitochondrial import site proteins ISP42 and
RT WOMB3.";
RL Biochim. Biophys. Acta 1153:345-347(1993).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
(Potential).
CC -1- SIMILARITY: TO E. COLI YFAL.
CC -1- SIMILARITY: SOME, TO FUNGAL MITOCHONDRIAL IMPORT SITE PROTEINS
ISP42 AND WOMB3.
CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
FRAMESHIFT IN POSITION 653.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000248; AAC74583.1; -.

```

```

DR EMBL; D90793; BAA15190.1; ALT_INIT.
DR EMBL; D90794; BAA15197.1; ALT_INIT.
DR EMBL; X73295; CAA51730.1; ALT_FRAME.
DR PIR; S34315; S34315.
DR EcGene; EC11780; ydek.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hypothetical protein; Membrane; Lipoprotein; Signal;
KW Complete proteome.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 1 1325 HYPOTHETICAL_LIPOPROTEIN_YDEK.
FT LIPID 19 19 N-ACYL DIGLUTERIDE (POTENTIAL).
FT CONFLICT 884 884 N->K (IN REF. 3).
FT CONFLICT 1317 1317 M->S (IN REF. 3).
SQ SEQUENCE 1325 AA; 136514 MW; 26A3A06FA19AD7D CRC64;

alignment_scores:
Quality: 236.00 Length: 961
Ratio: 0.524 Gaps: 50
Percent Similarity: 46.826 Percent Identity: 20.604

alignment_block:
US-09-303-518D-651 x YDEK_ECOLI ..
Align seg 1/1 to: YDEK_ECOLI from: 1 to: 1325

634 GATTATTCCTACTCCGCGCATGGTTAATTGGC.....GGCAA 671
77 AspValaIatYrSP...AlaTrYleuValGlyTrpTrYglyThrIolYva 92
672 TACACATATGACAGGTTGGGAAATATATGCGTANTTAGTTGAGCGCG 721
92 LLeuAsnIleLeuAlaGlyLysAlaSerLeuThrIleThrThrs 109
722 ATGTGCGCCATGCCAAGCAGTATGGCCGTATGCCGATGGAGTGGCGCA 771
109 erValIleGlyAlaAsnGlu.....AspSerGluGlyThrVal 121
772 GCGCAGCAGCGTTCCGCAATGTTATTTATGACAAACAAACATTAATG 821
122 AsnValleuGlyGlyThrTrpArgLeuTrYAspSerGlyAsnAsnAlar 138
822 GCTGCTCAACGAGTTTTCACAAACCGGTCACCTTATTCGCGCAGGAAA 871
138 gProLeuAsn...ValGlyGlnSerGly..... 146
872 ACGGTTTCAGCTGATACGCAAGATTGGTTTACGATGACATTTCACA 921
146 ..... 146
922 GCGCATACATACCGTCTNTTTGAACCGCGCAGTAAC.....GG 962
147 ....ThreGlyThrLeuAsnIleLysGlnLysGlyHisValAspLygl 161
963 ACATTTTTCCTTACATCAACAACAACGAGTACGGGTACGTAACAGAAA 1012
161 yTyValleuArgLeuLysSerSerThrGlyValGlyThrValAsnValG 178
1013 CCAACGAAAGGTTTCATCAACAAGCTTAAGTACAGACAGTCGACAG 1062
178 LuGlyGluAsp.....SerValleuThrThrGluLeu 188
1063 TTTCAGC.....GATCTTTGAATGAACGATTAAGA 1094
189 PheGlnIleGlySerTrYglnThrGlySerLeuAsnIleThrAspLygl 205
1095 ACCAGTT.....TACGCGCAGGCGGCTG 1117
205 yTyValThrSerSerIleValAlaIleLeuGlyTrYglnAlaGly.... 220
1118 TTAATCAGTACCGTCCACAGGTTAAACACAGCGTAAACCTTTTATC 1167
221 .....SerAsnGlyGlnValVal.....Val 227

```

1168 GATTACGCGCAACGCAAACTCATCTTATGCAACACATC..... 1206  
::: |||::: |||::: |||:::  
228 GluLysGlyGlyLysIleThrPheLysAsnAsnAspSerIleGluPhe 244  
1207 ..... ACCAAGCGCGCGGCGTTGTATTGTAAGGTGATTGTA 1246  
|||::: |||::: |||:::  
244 eGlnIleGlyAsnGlnGlyThrIle..... 252  
1247 CGGTCGCGCTGAAACCAACGTAAGCGCGGCGGCTTCATATC 1296  
::: |||::: |||::: |||:::  
253 ..... GlnAlaThrIleArgGluGlyGlyLeuValThr 263  
1297 AGTGAAGACAGTACCGTTACTTGAAAGTAACGGCGTGCAACGACGC 1346  
::: |||::: |||::: |||:::  
264 AlaGlnAsnThrIleIleGlyGlyAsnAlaThrGlyLe..... 276  
1347 CCTGTCCAAATGCGCAAGGCGCGCTGCACGTTCAAGCCAAAGGGGAAA 1396  
|||::: |||::: |||:::  
277 ..... GlyThrLeuAsnValGln..... AspG 284  
1397 ACCAAGCGCTGATCAGCGTG..... GCGCAG 1422  
::: |||::: |||::: |||:::  
284 LnsAspSerValIleThrValArgArgLeuThrAsnGlyThrPheGlyAsn 300  
1423 GGTACAGTCATTTTGGATCAGACGAGACGATTAAGGCAAAACACAGC 1472  
|||::: |||::: |||:::  
301 GlyThrValAsnIleSerAsnAsnGlyLeuIleAsnAsnLysGlu..... 315  
1473 CTTTACTGAAATGCGCTGTCAGCGGC... AGGGTACGGTGCAACTGA 1519  
::: |||::: |||::: |||:::  
316 .TyrSerLeuValGlyValGlnAspLysSerHisGlyValValAsnValT 332  
1520 ATGCCGATAATCAGTCAAC..... CCGCAGCAA 1548  
::: |||::: |||::: |||:::  
332 hTrsPlySglYHisTrpAsnPheLeuGlyThrGlyGluAlaPheAlaGlyTr 348  
1549 CTCTATTTCGGCTTCGCGGC... GACGTTGGATTTAACGGCGCATTC 1595  
::: |||::: |||::: |||:::  
349 IleThrIleGlyAspAlaGlyAspGlyGluLeuAsnValSerSerGluG 365  
1596 GCTTGTTCACCGTATTCAAATACGATGTAAGGGCGGATGTCGTCG 1645  
::: |||::: |||::: |||:::  
365 Y..... LysValAspSerGlyIleIleThrAlaG 375  
1646 ATCATATGCCACACACATCCACCGTTACCTACAGGGAATGAAGT 1695  
::: |||::: |||::: |||:::  
375 LysMetLysGluThrGlyThrGlyAsnIleThrValLysAspLysAsnSer 391  
1696 ATTACACACCGAGTGTAGAATATCATAGACTTAATTACGCAAGA 1745  
::: |||::: |||::: |||:::  
392 ValIleThrAsnLeuGlyThrAsn..... 399  
1746 AATTGCTACACGGTTGGTGGCAGAAAGATACACCAAAACGACAG 1795  
::: |||::: |||::: |||:::  
400 .LeuGlyTyrAspGly..... HisG 406  
1796 GCGCGCTCAACTTGTATTACACCGCGCAGACAGACCGACCGCTG 1845  
|||::: |||::: |||:::  
406 LysLeuMetAsnIleSerAsnGlnGlyLeuVal..... Val 417  
1846 CTTTCGCGGGAACAATTTAATC..... GGCACAT 1877  
::: |||::: |||::: |||:::  
418 SerAsnGlySerSerLeuGlyTyrGlyGluThrGlyValGlyAsnVal 434  
1878 CAGGCAAAACAGCAACTGTTTTCAGCGGCAACCGACACCGCAGC 1927  
::: |||::: |||::: |||:::  
434 IleThrIleThrGlyLysMetTrp..... GluValAsnLysAsnVal 448  
1928 CCTACATATTTAGACAGCGGCTGCTCAAAATAGAGATATCCACAA 1977  
::: |||::: |||::: |||:::  
448 alThrThrThrIleGly..... ValAlaGlyVal..... 457

1978 GGAGAAATCTGTGGACACACGACTGATCCAGCATGTTAAAGCGGA 2027  
458 ..... G1 458  
2028 AATTTCCATATTCAGGCGGCGGCGGATTTCCGCAATGTGCCA 2077  
|||::: |||::: |||:::  
458 YAsnLeuAsnIleSerAspGlyGlyLysPheValSerGlnAsnIleThrP 475  
2078 AAGTGAAGCGGATTCATTTGAGCAATCAGCCAGCAGTTTGGT 2127  
::: |||::: |||::: |||:::  
475 helen... GlyAsp..... 478  
2128 GTGCGCCGATCAAGCCATACATCTGACGTTGCGACTGACGACNG 2177  
479 ..... LysAlaSerGlyIleG1 484  
2178 TCTGCAATTTGTGCGAANAANCATTACGAGATTAAGTATGCTT 2227  
|||::: |||::: |||:::  
484 YThrLeuAsnLeuMetAspAlaThrSerSerPheAspThrValGlyLea 501  
2228 CATTCACTAAGACGACGACGTCAGTGTACGTCGTCGACGTCGTCG 2277  
::: |||::: |||::: |||:::  
501 snValGlyAsnPhe... GlySerGlyIleValAsnValSerAsnGlyAla 516  
2278 NNTTNAANCTCNCNGGCGTCNNCACTNANCGCAATCTTAGTCAAA 2327  
::: |||::: |||::: |||:::  
517 ThrLeuAsnSerThrGlyThrGlyPheIleGlyLysAlaSerGlyLys 533  
2328 TGCC..... GATPACACGTTAT..... ACAGTCA 2350  
|||::: |||::: |||:::  
533 sGlyIleValAsnIleSerThrAspSerLeuThrPAsnLeuLysThrSers 550  
2351 GCCACAACGCC..... ACCCAAAACGCC 2373  
|||::: |||::: |||:::  
550 erThrAsnAlaGlnLeuLeuGlnValGlyValLeuGlyThrGlyGluLeu 566  
2374 AACCTTACCTCGTGGCAATGCCAAGACACATTTATCAAGCCACATT 2423  
|||::: |||::: |||:::  
567 AsnIleThrThrGlyGlyIleValAlaLysAlaArgAspThrGlnIleAla 583  
2424 AAC..... GGCACNCATCGGNTTGGGCAATGCTCA... 2457  
|||::: |||::: |||:::  
583 uAsnAspLysSerLysGlyAspValArgValAspGlyLysAsnSerLeu 600  
2458 ..... TTTAATCTAAGCAACAGCGCCGACAAAACGGCAGTCTGAC 2499  
|||::: |||::: |||:::  
600 euGluThrPheAsnMetArgValGlyThrSerGlyThrGlyThrLeuThr 616  
2500 CTTTCGCAACGCTAAGCAACGTA..... 2526  
|||::: |||::: |||:::  
617 LeuThrAsnAsnGlyThrLeuAsnValGlnGlyGluValThrLeuG1 633  
2527 ..... AGCCAT... 2532  
633 yValPheGluProAlaValGlyThrLeuAsnIleGlyAlaAlaHisGlyG 650  
2533 ..... TCCGACCTCAACGCAATGTCTCCCTAGCCGATAGCA... 2571  
|||::: |||::: |||:::  
650 lValAlaAlaLysPalaGlyPheIleThrAsnAlaThrLysValGluPhe 666  
2572 ..... GTATTCCATTTTGAACACAGCCGCTTTAC 2600  
|||::: |||::: |||:::  
667 GlyLeuGlyGluGlyValPheValPheAsnHisThrAsnSerAspAl 683  
2601 CGGACAA..... CTCACGGCGACAGAGANACAGCATTTAC 2635  
|||::: |||::: |||:::  
683 agLysThrGlnValAspMetLeuIleThrGlyAspAspLysAspLysVal 700  
2636 ACTTAAAGACAGCAATGACGCTG..... CGTCAAGGACGACGAATTA 2679  
::: |||::: |||::: |||:::  
700 alIleHisAspAlaGlyHisThrValPheAsnAlaGlyAsnThrLysSer 716  
2680 GGCATTTTAACTTGACACGCGCACCATTAACCTCAATTCGCGCTATCG 2729



```

FT CONFLICT 985 985 N -> K (IN REF. 1)
FT CONFLICT 1009 1009 L -> S (IN REF. 1)
FT CONFLICT 1013 1013 Y -> S (IN REF. 1)
FT CONFLICT 1182 1182 K -> O (IN REF. 1)
FT CONFLICT 1314 1314 N -> Y (IN REF. 4)
FT CONFLICT 1451 1451 H -> N (IN REF. 1)
FT CONFLICT 1624 1624 E -> D (IN REF. 1)
FT CONFLICT 1628 1628 G -> G (IN REF. 1)
FT CONFLICT 1872 1872 A -> V (IN REF. 1)
FT CONFLICT 1875 1875 T -> P (IN REF. 1)
FT CONFLICT 1878 1878 MS -> LP (IN REF. 1)
FT CONFLICT 1936 1936 E -> A (IN REF. 1)
FT CONFLICT 1965 1965 MTADLP -> TTPPLS (IN REF. 1)
FT CONFLICT 1997 1997 G -> R (IN REF. 1)
SQ SEQUENCE 2021 AA: 203328 MW: 327FC42D7CB24668 CRC64;

```

alignment\_scores:

Quality:	234.00	Length:	1401
Ratio:	0.363	Gaps:	65
Percent Similarity:	45.967	Percent Identity:	19.914

## alignment\_block:

US-09-303-518D-651 x OMPA\_RICCN ..

Align seg 1/1 to: OMPA\_RICCN from: 1 to: 2021

```

622 GACAAACAGCGGATTTACTACTCCGCGCATGTTATATGGCGCA 671
||||| : : : : : : : : : : : : : : : : : : : :
319 ASPASNAlaSerAlaValThrPheThrAsnProValValThrGlyAl 335
672 TACACATATGACAGGTTGGGAAATPAATGCGTANTTATGTTAGCGG 721
||||| : : : : : : : : : : : : : : : : : : : :
335 AlAspAsnThrGlyAsnAlaAsnAlaValThrPheThrGlyAl 352
722 AT.....GTGGCCATGCCAAGCACTATGCGCCCTATG 753
|| : : : : : : : : : : : : : : : : : : : :
352 SPSerThrValThrGlyAsnIleGlyAsnThrAsnAlaLeuAlaThrIle 368
754 CGGATTCAGGTCGCGCAGCAGCGATTCGCCAATGTT..... 795
||||| : : : : : : : : : : : : : : : : : : : :
369 SerValGlyAlaGlyLysAlaThrLeuGlyAlaIleIleLysAlaThr 385
796 .....ATTATGACAAACAAACAAATPAATGCTGCTCAAGCAG 835
|| : : : : : : : : : : : : : : : : : : : :
385 rThrThrLysLeuThrAspAsnAlaSerAlaValThrPheThrAsnPro 402
836 TTTTCAAAACCGGC...TACCCTATTCGCGCAGGAAACGGTTTCAG 882
||||| : : : : : : : : : : : : : : : : : : : :
402 AlValValThrGlyAlaIleAspAsnThrGlyAsnAlaAsn..... 415
883 CTGATACGCAAGATTGTTTACGATGACATTTACAGAGCGCATACACA 932
415 ..... 415
933 TACCGCTTTTGAACCGCGCAGTACGACATTTTCCTTACATCCA 982
||||| : : : : : : : : : : : : : : : : : : : :
416 .....AsnGlyIleValThrPheThrGlyAl 424
983 ACAACAGCGTACGGGTACGGTAAACAAAC..... 1017
424 SPSerThrValThrGlyAsnIleGlyAsnThrAsnAlaLeuAlaThrIle 440
1018 .....GAAAGSTNTCCAAATCCAAAGCTTAAAGTACA 1049
441 SerValGlyAlaGlyLysAlaThrLeuGlyAlaIleIleLysAlaThr 457
1050 GACAGTCGCACTGTTGAAGATCTTTGAATGAACATGATTAAGACGAG 1099
||||| : : : : : : : : : : : : : : : : : : : :
457 rThrThrLysLeuThrAspAsnAlaSerAlaValThrPheThrAsnPro 474
1100 TTTACGGCGGAGGGGTGTATATGATGACGTCGAAGTTAAACAGCGT 1149
|| : : : : : : : : : : : : : : : : : : : :

```

```

474 AlValValThrGlyAlaIleAsp.....AsnThrGly 484
1150 GAA...AACCTTCCTTTATGATTAAGGACGACGCAACTCAATCTATC 1196
||| : : : : : : : : : : : : : : : : : : : :
485 AsnAlaAsnGlyIleValThrPheThrGlyAspSerThrValThrGly 501
1197 AAC.....AACATCAACCAAGCGCGCGC 1222
501 yAsnIleGlyAsnThrAsnAlaLeuAlaThrIleSerValGlyAlaGlyL 518
1223 GTTGTATTTTGAAGTCGATTTTACGCTTCGCCCTGAACAAACGAAACG 1272
518 ySalatThrLeuGlyAlaIleIleLysAlaThrThrThrLysLeuThr 534
1273 TGCGACAGCGCGGCGCTTCATATCAGAGACAGTACGCTTCTGGAA 1322
535 ASPASNAlaSerAlaValThrPheThrAsnProValValThrGlyAl 551
1323 AGTAAACGGCGTGGCAACGACGCGCTGTCCAAATTCGCAAGGACGCGC 1372
551 AlAspAsnThrGlyAsn..... 557
1373 TGACGTTCAAGCCAAAGGGGAAACCAAGCTCGATCAGC...GTGGCC 1419
558 .....AlaAsnAsnGlyIleValThrPheThrGly 567
1420 GACGCTACAGTCATTTTGTGATCAGCAGCAGATPAATGAACAAACAA 1469
||| : : : : : : : : : : : : : : : : : : : :
568 ASPSerThrVal.....ThrGlyAsnIleGlyAsnThrAs 579
1470 AGCCTTTAGTGAATCGGCTTGNTCAGCGGCGAGGGGTACGTCGCAACTGA 1519
579 nAlaLeuAlaThrIleSerValGlyAlaGlyLysAlaThrLeu..... 593
1520 ATGCCGATTAATCACTTAACCCGACAAACTATTTTCGCTTCGCGGC 1569
594 .....Gly 594
1570 GGA.....CGTTGATTTTAAACGGCGCATTC 1595
||| : : : : : : : : : : : : : : : : : : : :
595 GlyAlaIleIleLysAlaThrThrThrLysLeuThrAspAsnAlaSerAl 611
1596 GCTTCGTTCCACCGTATTCAAATACCGAT.....GAAAGGG 1633
611 AlValThrPhe.....ThrAsnProValValThrGlyAl 623
1634 CGATGATTGCMATCATATGCCACAAACACATCCACCGTTACCATTAACA 1683
||||| : : : : : : : : : : : : : : : : : : : :
623 lAlIleAspAsnThrGlyAsnAla...AsnAsnGlyIleValThrPheThr 638
1684 GGGATGAAGATTAATACAAACGAGTGTGAAGAATATCAATGACTT... 1731
639 GlyAsnSerThrValThrGlyAsnIleGly...AsnThrAsnAlaLeuAl 654
1732 .....AATTACGCAAGAAATTCGC...TCAACGCTGTTGGCG 1771
654 AlThrValAsnValGlyAlaGlyLysAlaThrLeuGlyAlaValIleL 671
1772 AGAAAGATACGACCAAA...ACGAACGGG..... 1797
671 ySalatThrThrThrLysLeuThrAsnAlaIleAspAsnThrThrGlyValAs 687
1798 CGGCTCACCTGTTTACGACGCGCGGAGAGACGCGCACCGCGTGCCT 1847
688 AsnValAsnAlaValLeuThrGlyAlaIleAspAsnThrThrGlyValAs 704
1848 TTCGGCGGGAACA...AATTAAACGCAACATCAGCAACAAACAGGCA 1894
704 pAsnValGlyValLeuAsnLeuAsnGlyAlaLeuSerGlnValThrGlyAl 721
1895 AACTGTTTTCACGCGCAGACGACACCGCCTACATATCATTTAGCA 1944
721 snIle.....GlyAsnThrAsnAlaLeuAlaThrIleSerValGly 734

```



```

|||||
1302 LeuAspPheAsnAlaLysAsnGlyThrValThrLeuAsnAsnValAs 1318
1320 CGTTGGCGAAGCGCCCGCGGCGGAAATGTCGCATTAATCC 3318
1318 nVAlAlaGlyThrValLysAsnThrGlyThrAsnAsnGlyThrLeuI 1335
1319 A.....GGCGAGGAGAGAGAA 3335
1335 LeValLeuGlyAlaSerAsnLeuAsnArgValAsnGlyLeuAlaMetLeu 1351
1336 AAACGGGTGGCAGC.....GGATAAGACAGCGCCT 3367
1352 LysValGlyAlaGlyAsnValThrIleAlaLysGlyGlyAsnValLysI 1368
1368 TGGCGAAGAGCGGAGCGGAAAC.....CCGGCGCGCT 3402
1368 eGlyGlyIleGlnGlyThrGlyThrAsnThrLeuThrLeuProAla...H 1384
1368 ACCACGCGCTCCCGCGCGCGCGCGCGCGCGGATTTGCC..... 3446
1384 IsPheLysLeuThrGlySerIleAsnLysThrGlyGlyAlaLeuLys 1400
1347 .....GCAACCGACGCGCCACCGCACTCAAC 3475
1401 LeuAsnPheMetAsnGlySerValSerGlyValAlaGlyThrAlaAl 1417
1417 CCCACCGAGCGGACGCTGATGATGACGCTTATGCGCAATGCGGTTGAGT 3525
1417 aaAsnSerValGly.AsplIeThrThrAlaGlyAlaThrSerPheAlaSer 1433
1417 GAATTTCCGCGCACGCTCAACAGCTTTTCCGCGTACAGAGCAATGGA 3575
1434 SerValAsnAla.LysGlyThrAlaThrLeuGlyGlyThrThrSerPhe 1450
1434 CGCGGTGTTGGCGAGCGCGCGCGCGCGCTT.....GGACAAGN 3619
1450 IaIstThrPheThrAsnThrGlyAlaValThrLeuAlaLysIleSerIle 1466
1450 GCATCCGCGNACCAACACTACCTGCTGCAAGATTTCCGCGCTACCGC 3669
1467 ThrSerPheAlaLysAsnValThrAlaThrSerPheValAlaAsnSerI 1483
1467 C.....ACAAACGACCTGCGCGCAATCGTATGACAGA 3704
1483 aThrIleAsnPheGlyAsnSerLeuAlaPheAsnSerAsnIleThrGlys 1500
1483 AA.....ACCTCGAGCGGCGCGGCGGATCTCTTTTGGCACAAC 3748
1500 eGlyThrThrLeuThrLeuGlyAlaAsnGlnValThrThrGlyThr 1516
1500 GACGCAAAACANCTTCGACGAGCGGATGCGCACTCGCGCAGCGCTTGGC 3798
1517 GlySerPheThrAspThrLeuThrLeuAsnThrThr..... 1528
1517 CACGCGCGGCTTTCGCGCAATACGCGATCGCAGGTTTCGACATCGCAT 3848
1529 .....PheAspGlyAlaAlaLysSerGlyLysn..... 1538
1529 CAGCAGCGGCGGCTTTAGCAGCGGCACTTNTCAGACGCGCATCGAG 3898
1539 .....IleLeuIleLysSerGlySerThrLeuAspLeuSer 1550
1539 GCAAAATCCGCGCGCGGCGGCTGCAATGCGCATTCAGCGCAGCATACGC 3948
1551 GlyValSerAsnLeuAlaLeuValAlaThrAlaThrAsnPheAsnMetAs 1567
1551 GCCGGTTTCGCGGATTCGCGATCGAACCGTACATCGCGCAGCGCT 3997
1567 nAsnIleSerProAspThrLysTyrThrValIleSerAlaGluThrAla 1583
seq_name: SwissProt_40:PERT_BORPE

```

```

seq_documentation_block:
ID PERT_BORPE STANDARD: PRT; 910 AA.
AC P14283;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Pertactin precursor (Outer membrane protein P.69) (P.93).
GN PRN OR OMP69A.
OS Bordetella pertussis.
OC Bacteria: Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-ISOLATE CN2992.
RX MEDLINE=89264462; PubMed=2542937;
RA Charles I.G., Dougan G., Pickard D., Chatfield S., Smith M.,
RA Novotny P., Morrissey P., Fairweather N.F.;
RT "Molecular cloning and characterization of protective outer membrane
RT protein P.69 from Bordetella pertussis.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:3554-3558(1989).
[2]
RP REVISIONS TO 264 AND 332.
RX MEDLINE=92407514; PubMed=1527510;
RA Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;
RT "Cloning, nucleotide sequence and heterologous expression of the
RT protective outer-membrane protein P.68 pertactin from Bordetella
RT bronchiseptica.";
RL J. Gen. Microbiol. 138:1697-1705(1992).
[3]
RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=96196517; PubMed=8609988;
RA Emsley P., Charles I.G., Fairweather N.F., Isaacs N.W.;
RT "Structure of Bordetella pertussis virulence factor P.69 pertactin.";
RL Nature 381:90-92(1996).
CC -1- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STAGE OF WHOOPING COUGH.
CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)
CC CONCENTRATION.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J04560; AAA22980.1; ALT_SEQ.
DR PIR: A32560; A32560.
DR InterPro: IPR003992; pertactin.
DR InterPro: IPR003991; pertactin-related.
DR PRINTS: PR01482; PERTACTIN.
DR PRINTS: PR01484; PERTACTINFAMILY.
KW Outer membrane; Signal; Virulence; Repeat.
FT SIGNAL 1 34
FT CHAIN 35 910 P.93.
FT CHAIN 35 711 PERTACTIN (P.69).
FT PROPEP 712 910 POTENTIAL.
FT SITE 260 262 CELL ATTACHMENT SITE (INVOLVED IN
FT ADHESION TO VARIOUS EUKARYOTIC CELL
FT LINES).
FT DOMAIN 266 290 5 X 5 AA TANDEM REPEATS OF G-G-A-V-P.
FT REPEAT 266 270 1.
FT REPEAT 271 275 2.
FT REPEAT 276 280 3.
FT REPEAT 281 285 4 (APPROXIMATE).
FT REPEAT 286 290 5 (APPROXIMATE).
FT DOMAIN 579 593 5 X 3 AA TANDEM REPEATS OF P-Q-P.

```



```

460 .....:::||||:
2820 TTGGTAGATCCGTTTCAACACGCTGACGGTAAACGGCAATGACN 2869
      ::::: |||||:::|||||
465 agluAa...GlyArgPheLysValLeuThrValAsn...ThrLeuAlaG 480
2870 GTCAAGACATTCGCTTATATGTCGAACTCTTCGGCTACCGAAGCGC 2919
      ::||| |||||:::
480 LysSerLysLeuPheArgMetAsnValPheAlaAspLeuLysLeuSerAsp 496
2920 AAATTAAGCTGGCGGAAAGTTCCGAGGNACTTACACCTTGCGCTCA 2969
      |||||::: |||||::: |||||::: |||||::: |||||:::
497 LysLeuValValMetGlnAspLysSerLysGlnHisArgLeuThrPala 513
2970 CAATACCGGCAAGACCCGTAAGCTCGATCATGACGTTAGGTAAG 3019
      ::|||::: |||||::: |||||::: |||||::: |||||:::
513 gAsnSerLysSerLysProLysSerLysAsnThrLeuLeuValGln 529
3020 GGAAGACAAACACCGCTGTCCGAAACCTTAATTACCTGCAAC 3069
      |||||::: |||||::: |||||::: |||||::: |||||:::
530 .....ThrProLeuGlySerAlaAlaThrPheThrLeuAlaAsn 542
3070 GAA.....CACGTGATCGCGCGCGCTGCTTACCACTCATCCGAA 3113
      ::||| ||||| |||||::: |||||::: |||||::: |||||:::
543 LysAspGlyLysValAspLysLysGlyThrLysArgLysArgLeu..... 556
3114 AGACGGCGACTTCGCGCTGATATCCGGTCAAGACAGAGCTTCCG 3163
556 ..... 556
3164 ACAAACTCGGCAAGCAGAAAGCAAAAGCGGAAAGACAAACACGCG 3213
3214 CAAGGCTTGACGCGCTGATTCGGCGCGCGGATCGCGGAAAGAC 3263
556 ..... 556
3264 AGAAACGTTTGGCGAAGCGCGCGCGCGGCGGGAATGTCGGA 3313
      |||||::: |||||::: |||||::: |||||::: |||||:::
557 .....AlaAlaAsnGlyAsnGlyGln.TripSerLeu 566
3314 TTATGACGCGGAGAGAAAGAAAGAGGGTGCAGCGGATTAAGACACG 3363
      ::|||
567 Val..... 567
3364 GCNTTGGCAAAACAGCGCGAAGCGGAAACCGCGGNTACACCGCTT 3413
      ::||| ||||| |||||::: |||||::: |||||::: |||||:::
568 .....GlyAlaLysAlaProLysProLysProLysProLysProL 580
3414 CCGCGCGCGCGCGCGCGCGCGGATTCGGCAAGCGGCAAGCGGCA 3462
      ::|||::: |||||::: |||||::: |||||::: |||||::: |||||:::
580 InProGly.Pro...GlnProPro.GlnProProGlnProGlnProGln 595
3463 ..CCGCAACCTCAACCCCAACCGCACCGGACCTGATNACCGCTATGCC 3510
      ||||| ||||| |||||::: |||||::: |||||::: |||||:::
595 LysProLysProGlnProProLysArgLysGlnLeu...SerAlaAlaAla 610
3511 AAATAGCGGTTTGAATTTTCCGCGACGCTCAACACGCTT.....TT 3554
      |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
611 AsnAlaAlaValAsnThrLysGlyValGlyLeuAlaSerThrLeuTrp 627
3555 CGCGCTACAGAGAGATTTGACCGCGCTGTTTGGCGAAGCGCGGCAAC 3603
      ::||| |||||::: |||||::: |||||::: |||||::: |||||:::
627 rAlaGlySerAsnAlaLeuSerLysArgLeuGlyGlyLeuValGlyLeuAsp 644
3604 .....GCGTTTGGACAAGCGCATCCGAGNACCAACACATAC 3642
      ::||| ||||| |||||::: |||||::: |||||::: |||||:::
644 roAspAlaGlyLysAlaTrpGlyArgGlyPheAlaGlnArgLysGlnLeu 660
3643 CGTTCCAAAGATTTCCGCGCTACCGCGCAACAAACCGACCTCGCCAA 3692
      ::||| |||||::: |||||::: |||||::: |||||::: |||||:::

```

```

661 AspaSnaArgAlaGlyArgArgPheAspGlnLysValAlaGlyPheGlu 677
3693 CGGTATGACGAAACACCTC.....GGCAGCGGCGCGGTGGCA 3730
      ::|||::: |||||::: |||||::: |||||::: |||||::: |||||:::
677 uGlyAlaAspHisAlaValAlaValAlaGlyLysArgTrpHisLysLysG 694
3731 TCCGTGTTTTCGACACCGACCGGCAACCAACCTTCGACGAGCGGATCG 3780
      |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
694 LysLeuAlaGlyLysTrpThrArgLysArgArgLysPheThrLysLysGly 710
3781 AACTCGGACGCGCTTCGCGCGCGCGCGCTTTCGCGCAATACGCAATCG 3830
      ::|||::: |||||::: |||||::: |||||::: |||||::: |||||:::
711 GlyHisThrLysSerValHis.....ValGlyLysLysLysLysLys 724
3831 CAGTTGACATTCGGCATTCAGCACGCGCGGCGGCTTTAGACGCGGCA 3880
      |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
724 rLeAlaAspSerGlyPheLysLysAspAlaThrLeuArgAlaSerLys 741
3881 TN.....TCAGACGCG.....ATCGAGCG 3900
      |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
741 euGlnAsnAspPheLysValAlaGlySerAspGlyLysAlaValLysGly 757
3901 AAAATCGCGCGCGCGCTGCTGATTCAGGCAATTCAGCAGCATACCGCG 3950
      ::||| ||||| |||||::: |||||::: |||||::: |||||:::
758 LysLysArgLysThr.....HisGlyValGlyLysSerLeuGlnAl 770
3951 CGGT.....TTGCGGAGATTCGGCATTCGCAACCGCATTCG 3985
      |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
770 agLysArgArgPheThrHisAlaAspGlyLysPheLeuGlnProGlnAla 787
3986 GCGCAACGCGCTATTCCTCCAAAGCGGATTCACCGCTACGAAAC... 4032
      ::|||::: |||||::: |||||::: |||||::: |||||::: |||||:::
787 LysLeuAlaValPheArgAlaGlyLysLysLysLysLysLysLysLys 803
4033 GTCAATATCGCCACCGCGCTTGCCTTCACCGGNTACCGGCGGCGCAT 4082
      ::|||::: |||||::: |||||::: |||||::: |||||::: |||||:::
804 LeuArgValArgAspGlyLysLysSerValLeuLysArgLeuGlyLe 820
4083 TAAGGAGATTTATTCATTCAAACCGGCGCAACACATTCATTCATCNCCT 4132
      ::|||::: |||||::: |||||::: |||||::: |||||::: |||||:::
820 uGlnValGlyLysArgLysLysLysLysLysLysLysLysLysLysLys 837
4133 ATTTNAGCGCTCTCTATACGAGTCCGCTTCGGGCAAGTCCGAAACGCG 4182
      ::|||::: |||||::: |||||::: |||||::: |||||::: |||||:::
837 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 853
4183 CTCATATCCGCGGATTTGCTCAGAGATTTGCGCAAAACCGCGAGTGGGA 4232
      ::|||::: |||||::: |||||::: |||||::: |||||::: |||||:::
854 ThrAsnGlyLysLysLysLysLysLysLysLysLysLysLysLysLys 869
4233 ATGGGCGGTAAACCGCAATCAAGGTTTCAGCGCTTCGTCGTCAGCGCTG 4282
      ::|||::: |||||::: |||||::: |||||::: |||||::: |||||:::
869 uLeuGly.....LeuGlyMetAlaAla 877
4283 CCGCGCGCAAGGCGGCAACGCGGAGCG.....CAACACAGCGCGGCG 4326
      ::|||::: |||||::: |||||::: |||||::: |||||::: |||||:::
877 LysLeuGlyArgGlyHisSerLeuLysLysLysLysLysLysLysLys 893
4327 ATCAATTTAGGCTACCGCTGG 4347
      ::|||::: |||||::: |||||::: |||||::: |||||::: |||||:::
894 ProLysLeuAlaMetProLys 900
seq_name: SwissProt_40:PERT_BORPA
seq_documentation_block:
ID PERT_BORPA STANDARD; PRT; 922 AA.
AC P24328;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, last sequence update)
DE 01-OCT-1996 (Rel. 34, last annotation update)
DE Pertactin precursor (Outer membrane protein P.70) (P.95).
GN PRN.
OS Bordetella parapertussis.

```



OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
OC Bordetella.  
OX NCBI\_TaxID=519;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ISOLATE CN2591;  
RA MEDLINE=91251711; PubMed=2041476;  
RA Li L.J., Dougan G., Novotny P., Charles I.G.;  
RT "P-70 pertactin, an outer-membrane protein from Bordetella  
parapertussis: cloning, nucleotide sequence and surface expression in  
Escherichia coli".  
RL Mol. Microbiol. 5:409-417(1991).  
CC -1- FUNCTION: AGGLOUTININ THAT BINDS TO EUKARYOTIC CELLS; A PROCESS  
CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN  
CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.  
CC -1- SUBUNIT: MONOMER.  
CC -1- SUBCELLULAR LOCATION: Outer membrane.  
CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.  
CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)  
CC CONCENTRATIONS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in. no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: X54547; CA38419.1; -;  
DR EMBL: A26124; CA01786.1; -;  
DR EMBL: A19182; CA01454.1; -;  
DR PIR: S15204; S15204.  
DR PIR: S14659; S14659.  
DR InterPro: IPR003992; pertactin.  
DR InterPro: IPR003991; pertactin-related.  
DR PRINTS: PRO1482; PERTACTIN.  
DR PRINTS: PRO1484; PERTACTINFAMLY.  
KW Outer membrane; signal; Virulence; Repeat.  
FT SIGNAL 1 34  
FT CHAIN 35 922  
FT CHAIN 35 647  
FT PROPEP 648 922  
FT SITE 260 262  
FT DOMAIN 266 290  
FT REPEAT 266 270  
FT REPEAT 271 275  
FT REPEAT 276 280  
FT REPEAT 281 285  
FT REPEAT 575 603  
FT DOMAIN 922 95178 MM; 3DF7BC58D4712478 CRC64;  
FO SEQUENCE  
-----  
alignment\_scores:  
Quality: 227.00 Length: 1119  
Ratio: 0.473 Gaps: 56  
Percent Similarity: 42.895 Percent Identity: 20.912  
-----  
alignment\_block:  
US-09-303-518d-651 x PERT\_BORPA ..  
Align seg 1/1 to: PERT\_BORPA from: 1 to: 922  
-----  
1279 GCGCGGGCGGTCATATCATGAGAGACGTACCTGTACTTGGAAGTAAA 1328  
|||||  
58 GYAlaGlyValArgThrAlaThrGlyThrThrIle.....LysValse 72  
1329 GCGCGGGGCAAGACGCGCTGTCCAAATGTGGCAAGGACGCTGCACG 1378  
|||||  
72 rGly.....A 74

1379 TTCAGCCAAAGG.....GAAACCAAGGCTG...ATCAGCTG 1416  
|||||  
74 rGlnAlaGlnGlyValLeuLeuGlnAsnProAlaIleLeuLeuArgPhe 90  
1417 GCGCGGCTACAGTCAATTTTGATCAGCAGGACGACGATTAAGCAAAA 1466  
:::|||||  
91 GlnAsnGlySerValThrSerSerGlyGlnLeuPheAspGlnGlyValAr 107  
1467 ACAAGCCTTAGGAATTCGGCTTGNTCAGCGCAGG..... 1503  
107 gArgPheLeuGlyThrValThrValLysAlaGlyLysLeuValAlaAspH 124  
1504 .....GCTAGCGTG 1512  
124 IsAlaThrLeuAlaAsnValSerAspThrArgAspAspArgLysLeAla 140  
1513 CAACGATATCCGATATATCATGTTCAAC.....CCGACAAACTGA 1553  
|||  
141 LeuTyValAlaGlyGlnGlnAlaGlnAlaSerIleAlaAspSerThrIle 157  
1554 TTTCGGCTTCGCGGCGGCTTGATTAACGGGCAATTCGTTGCT 1603  
|||  
157 uGlnGlyAlaGlyValArgValGlu..ArgGlyAlaAsnValThrV 173  
1604 TCCACCGTATTCAAATATCCGATGAAGGGGCGATGATGNCATCAAT 1653  
:::|||||  
173 aGlnArgSerThrIleValAspGlyGlyLeuHisIleGlyThr.LeuG 189  
1554 GCCACAAACATCCACCGCTTACCATTACAGGAATGAAGTATTCACA 1703  
|||  
189 nProLeuGlnProGlnAspLeuPro..... 197  
1704 ACCAGTGTAAGATATCATAGACTTAATTACAGCAAGAAATTCCT 1753  
|||  
198 .....Pro 198  
1754 ACAACGGTGTGTTGGCGAGAAAGATATACACCAAAAGGCGGCTC 1803  
::: |||  
199 SerArgValValLeu.....GlyAs 205  
1804 AACCTTGTTCACAGCCCGCCGAGAACCCGACCCGCTGC..... 1846  
|||  
205 pThrSerValThrAlaValProAlaSerGlyAlaProAlaIleValPheV 222  
1847 ..TTTCGGCGGAGCAATTTAAACGGCAATCAGCAACAAACAGCGCA 1894  
|||  
222 alPheGlyAlaAsnGlnLeu..... 228  
1895 AACGTGTTTCAGCGCGAGACCGACACCGACGCGCTCAATTCATTAGA 1944  
::: |||  
229 .....ThrValAspGlyGlyHisIleThr..... 236  
1945 AGCGGCTGTCAAATATGAGGTATCC.....C 1973  
|||||  
237 .....GlyGlyArgAlaIleValAlaIleValAlaIleValAlaIleVal 252  
1974 ACAAGGAGAAATCGTGTGGGACCAACGACGTGATCNCACGACGTTAAAG 2023  
||  
252 IsLeuGlnArgGala.ThrIleArgArgGlyAspLaPro..... 264  
2024 CGGAAATTCATATTCAGGCGGCGGCGGCGGATTCGCCCAATGTT 2073  
|||||  
265 .....AlaGlyGlyAlaValProGlyGlyA 273  
2074 GCCAAAGTGAAGGCGATGNCATTTGAGCAATACGCGCCCAAGCAGTTT 2123  
::: |||  
273 IaValProGlyGlyAlaVal.....ProGlyGlyPhe 283  
2124 .....TGGTGCGACCGCATCAAGCATATCATC 2155  
|||  
284 GlyProLeuLeuAspGlyTrp..... 290  
2156 GTACACGTTGCACTGACNGGTCGTGACAAATTTGTGCAANAANCATT 2205

```
291 .....TYGlyValAspValSerasp.SerThrValasp..... 301
2206 ACGGAGATTAAGTGAATGCTTCATGACTAGTAAGACAGACNTAAGCGGCAN 2255
302 .....LeuAlaGlnSerIleValGluAlaProGlnLeuVal 314
2256 TGTNAGNCTNNCCNATNAGCNTNNNTNAAMNCTCNCNGGCGNTGCNNCAC 2305
314 aaia..... 315
2306 TNANAGCAATCTTGTGCAATGGCGATACAGCTTATACAGTCAGCCAC 2355
316 .....IleArgAlaGlyArgGlyAlaArgValThrValSerGly 328
2356 AAC.....GCCACCCAAAGCGCAACCTTAGCGCTCGTGCGCAATGCCCA 2399
329 GlySerLeuSerAlaProHisGlyAsnValIleGluThrGlyGlyAla 345
2400 AGCAACATTTAAT.....CAAGCCACATTAACGGCA 2431
345 aatgatgphieproproAlaSerProLeuSerIleThrLeuGlnAlaG 362
2432 ACNCATCGGNTTCGGCAATGCT..... 2454
362 lYAlaArgAlaGlnGlyArgAlaLeuLeuTyArgValLeuProGluPro 378
2455 ....TCATTTAATCTAGCAACAACGCCGCAAAAGCGAGTCTGACGCT 2501
379 ValLysLeuThrLeuAlaGlyAlaGlnGlyGlnGlyAspIleValAl 395
2502 TTCGCAACACGCTTAAGCAACGTAAGCCATTCGCGACTCAACAGGC... 2547
395 aThrGluLeuProProIle.....ProGlyAlaSerSerGlyProL 409
2548 ..AATGTCCTCCAGCGGTAAGGCGATTCATCTTTGAAAAACAGCCGC 2595
409 euAspValAlaLeuAlaSerGlnAla.....Arg 418
2596 TTTACGGACAACCTCAGCGGACAGAGANACAGATTCACCTTAAGA 2645
419 TrpThrGly.....AlaThrArgAlaValAspSerLeuSerIleAspAs 433
2646 CACGGAATGAGCGCTGCGTCAAGGCAAGCAATTAAGCAATTAACCTTG 2695
433 .....AlaThrTrpValMetThrAspAsnSerValGlyAlaLeuArgLeu 449
2696 ACAAGCCCAACATTCACATCAATTCGCTATCGCCAGATGCTGCAGCG 2745
449 ..... 449
2746 GCGCAACCGGAGAGNGTGTACAGACGCGCGCGCTTCGCGCGTTCC 2795
450 AlaSerAspGlySerVal..... 455
2796 CCTATTATCCGTTACACCGCAACTTCGGTAGAATCCGTTTCAACAGCG 2845
456 .....AspPheGlnGlnProAlaGluAla...GlyArgPheLysValL 469
2846 TGACGGTAAGCGCAATTAAGACNGTCAAGCAACATTCGCGCTTATGTCG 2895
469 euMetValAsp...ThrLeuAlaGlySerGlyLeuPheArgMetAsnVal 484
2896 GAACCTTTCGGCTACCGAAGCAGCAAAATTGAAGCTGGCGGAAGTTCCGA 2945
485 PheAlaAspLeuGlyLeuSerAspLysLeuValValMetArgAspAlaSe 501
2946 AGGNACTTACACCTTGCGGCTCAACAATAACCGGCAACGACCGTAAAGCC 2995
501 rGlyGlnHisArgLeuTrpValArgAsnSerGlySerGluProAlaSerG 518
2996 TGGATCAATTAAGCGTACTGGAAGGAAAGACACAACCGCTGTCCGGA 3045
::: ::: :|||||:::
518 lYAsnThrMetLeuLeuValGln.....ThrProArgGlySer 530
3046 AACCTTAATTCACCTCGCAAAAGCAA.....CAGTCGATGCGCGCGC 3089
531 AlaAlaThrPheThrLeuAlaAsnLysAspGlyLysValAspIleGlyTh 547
3090 GTGGCGTTACCACTCAATCCGCAAA...GACGGGAGATTCGCGCTGCATA 3136
547 rTyArgTyArgLeuAlaAlaAsnGlyAsnGlyGlnTrpSerLeu.Val 563
3137 ATCCGGTCAAGAACAGAGAGCTTTCGCAACAATTCGCGCAAGCAAGAAC 3186
563 ..... 563
3187 AAAAAACAGCGGGAAGAACAGCGCAAAAGCTTGACGCGCTGATTGC 3236
563 ..... 563
3237 GGCAGGCGCGATGCGCGCAAGAAAGCAAGAGCGTTGCCGAACGCGCC 3286
564 ....GlyAlaLysAlaProProAlaProLysProAlaPro...GlnPro 577
3287 GCGNAGCAGCGCGGGAATGTCCGATTATGCAAGCGCGAGAAAGAGAA 3336
578 GlyProGln..... 580
3337 AAACGGGTGACGCGGATTAAGACAGCGCNTTGGCGAAACAGCGCGAAGC 3386
580 ..... 580
3387 GGAACCGCGCGGNTACACACCGCTTCCCGCGCGCGCGCGCGCC 3436
581 .....ProGlyProGlnProProGln.ProProGlnProProGlnProPr 595
3437 GGGATTGCGCGCAACG.....CAGCCCGCAA...CGGCAACCTCAA 3474
595 o.GlnProProGlnProProGlnArgGlnProGlnAlaProAlaProGln 611
3475 CCCCAACCGCAGCGGACCTGATNAGCGGTATGCAATAGCGGTTGAG 3524
612 ProProAlaGlyArgGlyLeu...SerAlaAlaAlaAsnAlaAlaValAs 627
3525 TGAATTTTCGCGCCAGCTCAACAGCGTT.....TTCGCGCTACAGGACG 3568
627 nThrGlyGlyValGlyLeuAlaSerThrLeuTrpTyAlaGlnSerAsnA 644
3569 AATTGGAACCGGCTGTTGCCGAAGACCGCGCGCAAC.....GCN 3606
644 lAlaSerLysArgLeuGlyGlnLeuArgLeuAsnProAspAlaGlyGly 660
3607 GTTTGACAAAGCNGCATCCGGMNACCAACCAACACTACCGTTGCAAGATT 3656
661 AlaTrpGlyArgGlyPheAlaGlnArgGlnGlnLeuAspAsnArgAlaG 677
3657 CCGCGCTTACCGCCACAACACGACCTGCGCCAAATCGTATGACAGAAA 3706
677 yArgArgPheAspGlnLysValAlaGlyPheGlnLeuGlyAlaAspHisA 694
3707 ACCTC.....GCGCAGCGGCGCGTTCGCGCATCTGTTTGCAC 3744
694 lValAlaValAlaGlyAlaGlyArgTrpHisLeuGlyGlyLeuAlaGlyTy 710
3745 AACCGGACCGCAAAACANTTCAGACGAGCATGCGCACTGCGACGAGCT 3794
711 ThrArgGlyAspArgGlyPheThrGlyAspGlyGlyLysHisThrAspSe 727
3795 TGCCACAGCGCGCGCTTTTCGGCAATAAGCGCATGCGCAGGTTGCACATCG 3844
727 rValHis.....ValGlyGlyTyArgAlaThrTyAlaAlaAsnSerG 741
3845 GCATCAGCAGCGCGCGGCTTTTACGACGCGCGCANTCN..... 3882
741 lPheTyTrLeuAspAlaThrLeuArgAlaSerArgLeuGlnAsnAspPhe 757
```

```

3883 .....TCAGACGCC.....ATCGAGAGCAAAATCGCGCGC 3914
758 LysValAlaIaGlySerAspGlyTyrAlaValAlaGlyGlyTyrAlaThr 774
3915 CGTGGCGCATTCAGCGCATTCAGCGCATTCAGCGCGCGCGCGCGCG 3961
774 sgIyValIglYvalSerLeuGluIaGlyArgArgPheAlaIaAspG 791
3962 GATTTCGGCATTCAGCGCATTCAGCGCGCGCGCGCGCGCGCGCGCG 4011
791 LyrTrpPheLeuGluProGlnAlaGlyLeuAlaValPheArgValIgl 807
4012 GCGGATTCAGCGCATTCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 4058
808 GAlaTyrArgAlaIaAsnGlyLeuArgValArgAspGluIglYse 824
4059 GTTCACACCGTACCGCGCGCGCGCATTCAGCGCATTCAGCGCGCG 4108
824 rSerValIleuGlyArgLeuGlyLeuGlyIaGlyLysArgIleGlu 841
4109 CGCACACATTCATCCATCCATCCATTCATTCATTCATTCATTCAT 4158
841 IeAlGlyArgGlnValGlnProTyrIleLysAlaSerValLeuGln 857
4159 GCTTCGGCGCAAGTCCAGACCGCGCATTCAGCGCGCGCGCGCGCG 4208
858 PheAspGlyAlaGlyThrValArgThrAsnGlyIleAlaIaHlaArg 874
4209 TTTCGGCAAAACCGCGCATTCAGCGCGCGCGCGCGCGCGCGCG 4258
874 uLeuArgIglYThrArg...AlaGluLeuGly..... 883
4259 GTTTCAGCGTGCCTCCATCCATCCATTCATTCATTCATTCATTC 4308
884 .....LeuGlyMetAlaIaIaIaIaLeuGlyArgGlyHisSerLeu 897
4309 GCG.....CAACACAGCGCGCGCGCATTCAGCGCGCGCGCG 4347
898 AlaSerTyrGluTyrSerLysGlyProLysLeuAlaMetProTyr 912
seq_name: SwissProt_40:AG43_ECOLI

```

```

RA Sivasubram S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=ML 308-225;
RA Henderson I.R., Owen P.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
[4]
RP PRELIMINARY SEQUENCE OF 53-78.
RC STRAIN=ML 308-225;
RX MEDLINE=9291704; PubMed=2661530;
RA Caffrey P., Owen P.;
RT "Purification and N-terminal sequence of the alpha subunit of antigen
RT 43, a unique protein complex associated with the outer membrane of
RT Escherichia coli K-12.";
RL J. Bacteriol. 171:3634-3640(1989).
[5]
RP SEQUENCE OF 53-63.
RC STRAIN=K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12.";
RL Electrophoresis 18:1259-1313(1997).
[6]
RP GENE NAME.
RX MEDLINE=97257509; PubMed=9103983;
RA Henderson I.R., Meehan M., Owen P.;
RT "Antigen 43, a phase-variable bipartite outer membrane protein,
RT determines colony morphology and autoaggregation in Escherichia coli
RT K-12.";
RL FEMS Microbiol. Lett. 149:115-120(1997).
CC -1- FUNCTION: CONTROLS COLONY FORM VARIATION AND AUTOAGGREGATION. MAY
CC -1- SUBUNIT: HETERODIMER OF THE PERIPHERAL MEMBRANE PROTEIN (ALPHA
CC CHAIN) ANCHORED TO THE INTEGRAL OUTER MEMBRANE PROTEIN (BETA
CC CHAIN).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE ASSOCIATED.
CC -1- SIMILARITY: TO ADHESIN AIDA-I AND TO BORDERELLA PERTACTIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AE000291; AAC75061.1; ALT_INIT.
DR EMBL: D90838; BAA15825.1; ALT_INIT.
DR EMBL: D90839; BAA15832.1; ALT_INIT.
DR EMBL: U24429; AAB47869.1; -.
DR Ecocore: EG12686; flu.
KW Outer membrane; Signal; Complete proteome.
FT SIGNAL 1 52
FT CHAIN 53 551
FT STRAIN 552 1039
FT VARIANT 2 2
FT VARIANT 41 42
FT VARIANT 46 46
FT VARIANT 157 157
FT VARIANT 188 188
FT VARIANT 303 305
FT VARIANT 320 372
FT VARIANT 372 372
FT VARIANT 493 493
FT VARIANT 497 497
FT VARIANT 555 585
FT VARIANT 709 709
FT VARIANT 721 721
FT VARIANT 721 721

```

FT VARIANT 751 753 GHL -> SHE (IN STRAIN ML 308-225).  
 FT VARIANT 803 803 S -> P (IN STRAIN ML 308-225).  
 FT VARIANT 815 815 A -> V (IN STRAIN ML 308-225).  
 FT VARIANT 824 824 C -> S (IN STRAIN ML 308-225).  
 FT VARIANT 829 835 LNLVHTS -> MNLVHTS (IN STRAIN ML 308-225).  
 FT VARIANT 845 847 OCT -> LGA (IN STRAIN ML 308-225).  
 FT VARIANT 855 855 S -> T (IN STRAIN ML 308-225).  
 FT VARIANT 868 868 Q -> L (IN STRAIN ML 308-225).  
 FT VARIANT 1025 1025 S -> I (IN STRAIN ML 308-225).  
 FT CONFLICT 61 63 ETV -> TTT (IN REF. 5).  
 SO SEQUENCE 1039 AA; 106841 MM; 5170647C8DBEBB0 CRC64;

alignment\_scores:  
 Quality: 224.50 Length: 1240  
 Ratio: 0.409 Gaps: 55  
 Percent Similarity: 44.274 Percent Identity: 19.435

## alignment\_block:

US-09-303-518D-651 x AG43\_ECOLI ..

Align seg 1/1 to: AG43\_ECOLI from: 1 to: 1039

1072 TCTTGAATGAACGATATAAGAACAGTTTACGCGCGGCGGTGTTAA 1121  
 |||||  
 41 SerLeuAlaIaValThrSerLeuProValLeuAlaIaAspIleValVa 57  
 1122 TCAGTACCGTCCAAAGTTAAACAAACGTTGAAACCTTTCTTTATCGATT 1171  
 ::|||  
 57 IHisProGlyGluThrValaIaSnGlyGly..... 66  
 1172 ACGCAAGCGCAAACTCATCTATCAACACATCAACCA..... 1212  
 |||||  
 67 .....ThreLeuAlaSnHIsAspAsnGlnIleValaPhe 77  
 1213 ...GGCGGGCGGCTTTGTATTTTGAAGTATTTACGCTCGCCTGA 1259  
 ::|||  
 78 GlyThrThrAsnGlyMetThrIleSerThrGlyLeuGluTrpGlyProAs 94  
 1260 AACAACAAGACGCTGCAAGCGCGCGGCTCAT.....ATCATG 1300  
 ::|||  
 94 PAsnGlnIaAsnThrGlyGlyGlnTrpValaGlnAspGlyGlyThrAla 111  
 1301 AAGACATGACGCTTACT.....TGAAGATAAC.....GCG 1332  
 ::|||  
 111 snLyThrThrValThrSerGlyGlyLeuGlnArgValaIaSnProGlyGly 127  
 1333 GTGGCAAGACGCGCTGTCCAAAATCGCAAAAGC..... 1368  
 ::|||  
 128 SerValSerAspThrValIleSerAlaGlyGlyGlnSerLeuGlnI 144  
 1369 .....ACGCTGACGTTCAAGCCAAAGGGGAAACCAAG 1402  
 |||||  
 144 yAlaGlnAlaValaIaSnThrThrLeuAsnGlyGlyGlnTrpMetHisGluG 161  
 1403 GCTGCATGACGCTGGC..... 1419  
 |||||  
 161 yAlaIleAlaThrGlyThrValIleAsnAspGlyGlyTrpGlnValaI 177  
 1420 ...GACGCTACGATTTTGGATCAGACGACGACGATTAAGCAAAA 1466  
 |||||  
 178 LysProGlyThrValaIaThrAspThrValaIaSnThrGlyAlaGluG 194  
 1467 ACAAGCCTTACTGAATCGGCTGNTCAGCGC..... 1500  
 ::|||  
 194 yGlyProAspAlaGluAsnGly...AspThrGlyGlnPheValaIaArgGly 210  
 1501 .....AGGGTACGGTGCACACTGAAT.....GCC 1524  
 |||||  
 210 spAlaValaIaArgThrThrIleAsnLysAsnGlyArgGlnIleValaIaArg 226  
 1525 GATAAATCACTCAACCCGACAAACTCTAATTTGGCTTTCGGCGGACG 1574

227 GluGlyThrAlaAsnThrThrValaValIaTr.....AlaGlyGlyAs 240  
 ::|||  
 1575 TTTGGATTTAAACGGGCAATCCCTTTCGTTCCACCGTATCAAAATACCG 1624  
 ::|||  
 240 pGlnThrValaIaHisGlyHisAlaLeu.....AspThrThrL 252  
 1625 ATGAAGCGGCGATGATTCGNCATTAATGACACAAACATTCACCGTT 1674  
 ::|||  
 252 euAsnGlyGlyTrpGlnTrpValaIleAsnGlyGlyThrAlaSerAspThr 268  
 1675 ACCATTACAGGAATGAAGATTAACACACCGAGTGTAGATATATCA 1724  
 ::|||  
 269 ValaIaSerAsp..... 273  
 1725 TAGACTTAATTACGCAAGAAATTCCTACACAGCGTTGTTGGCGAGA 1774  
 ::|||  
 274 .....GlyTrpGlnIleValaIaLysAsnGlyValaIaIaGlyA 286  
 1825 GCAGAAAGCCGACCCNCTGCTTCCGCGCGCAACAAATTTAAGCGCA 1874  
 ::|||  
 299 .....AlaGlyGlyThrAlaThrAsnValaIaTh 307  
 1875 CATTCACCAAAACAGCGCAAACTGTTTTCAGCGCGACACCGACCGC 1924  
 ::|||  
 307 rLeuLysGln...GlyGlyAlaLeuValThrSerThrAlaIaThrValT 323  
 1925 ACGCCTACAAATCATTTAGAACGCGGTGTCAAAAATGGAAGTATCCA 1974  
 ::|||  
 323 hrcGlyIleAsnArgLeuGlyAla...PheSerValaValGluGly...Lys 337  
 1975 CAAGGAATTCGTGTGGACAAACGACTGG.....ATCNACCGCAC 2015  
 ::|||  
 338 AlaAspAsnValaValLeuGluAsnGlyGlyArgLeuAspValaLeuThrI 354  
 2016 GTTTAAAGCGGAAATTTCCATATTCAGGCGGCGGCGGATTTCC 2065  
 ::|||  
 354 yHisThrAlaThrAsnThrArgValaIaSpAspGlyGlyThrLeuAspVala 371  
 2066 GCAT.....GTGCCAAATGCAAGCGCATTCNCAT 2097  
 |||||  
 371 rGAsnGlyGlyThrAlaThrThrValSerMetGlyAsnGlyGlyValaLeu 387  
 2098 TTGAGCAATCAGCCCAAGCAGTTTGTGTCGACCGCATCAAGCCA 2147  
 |||||  
 388 LeuAlaSpSerGlyAlaIaValaIaSerGly..... 397  
 2148 TACAATCTGACAGCTTCGAGCTGACGNGCTGTGACAAATTTGTGCAAN 2197  
 |||||  
 398 .....ThrArgSerAsp.....GlyLysAlaPheSerIleGlyG 409  
 2198 AAANCAATTACCGACGATTAAGTGATT..... 2223  
 ::|||  
 409 LysGlyGlnAlaAspAlaLeuMetLeuGlyLysSerPheThrLeu 425  
 2224 ...GCTCATTTGACTAGACGACGACGACGACGACGACGACGACGACG 2270  
 |||||  
 426 AsnAlaGlyAspThrAlaThrAspThrThrValaIaGlyGlyLeuThr 442  
 2271 TNAAGTNTTNAANCTCNCNGGCGCTGACCTTNAANGCAATCTTA 2320  
 ::|||  
 442 rAlaArgGlyGlyThrLeuAlaGlyThrThrLeu..... 454  
 2321 GTGCAAAATGGCGATACAGTTATACAGTTCAGCCACCAACGCCAAC 2370  
 ::|||  
 455 .....AsnAsnGlyAlaIleLeuThrLeuSerGlyGlyThrValaIaSn 469  
 2371 GGCACCTTGAAGCTC...GTGGGCAATGCCCAAGCAATTTATATACAG 2417  
 ::|||

470 AsptHrLeuThrIleArgGluValAspAlaLeuLeuGlnGlySerLe 486  
2418 CACATTAAAGCGCAACNCATCGGTTCCGGCAATCTTCATTATATCAA 2467  
||| |||||| :||| :||| :||| :||| :||| :||| :||| :|||  
486 uThrGlyAsnGlySerValGluYserGlySerGlyThrLeuThrValS 503  
GCACAAGGCC.....GCACAAAACGGCAGTGTG 2496  
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
503 eAsnThrThrLeuThrGlnLysAlaValAsnLeuAsnGlnGlyThrLeu 519  
ACGCTTTCGACAAAC..... 2511  
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
520 ThrLeuAsnAspSerThrValThrThraSpValIlealagInlaGlyTh 536  
2512 .GCTAAGCAACAGTAAGCCATTCCGCACTCAACGGC..... 2547  
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
536 rAlaLeuLysLeuThrGlySerThrValLeuAsnGlyAlaIleasPro 553  
2548 ..AATGTCCCTAGCCGATTAAGCAGATATC..... 2577  
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
553 hrAsnValThrLeuAlaSerGlyAlaThrTrpAsnIleProAspAsna 569  
2578 .....CATTT 2582  
570 ThrValGlnSerValValAspAspLeuSerHisAlaGlyGlnIleHis 586  
2583 TGAAGAAGCGCGCTTACCAGCACACTCAGCGGCGAGCAAGANACAGCAT 2632  
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
586 eThrSerThrArg...ThrGlyLysPhe..... 594  
2633 TACACTTAAGAACAGCAGGAATGACGCTGCCGTACAGCAGCAATTAGCG 2682  
595 .....ValProAlaThrLeuLysValLys 602  
2683 AATTAAACCTTGACACGCGACCATTAACATACTCCGCTATCGCCA 2732  
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
603 AsnLeuAsnGlyGlnAsnGlyThrIleSerLeuArg..... 614  
2733 CGATGCTGACAGGCGCGCAACCGCGCAGNGTGCAGACACCGCGCGCC 2782  
614 ..... 614  
2783 GTTGGCGCCGCTCCCTATTATCCGTTACACCGCCCACTTCGGTAGAATCC 2832  
615 .....ValArgProaspMetAlaGlnAsn 622  
2833 CGTTTCAACAGCGTGACGAGTAACGCGCAATTAACNGTCAAGACATTT 2882  
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
623 AsnAlaAspArgLeuValIleaspgly..... 631  
2883 CCGCTTATGTGCGAACTCTTCGGCTACCGAACGCAAAA.....TTGA 2926  
632 .....GlyArgAlaThrGlyLysThrIleLeuA 641  
2927 AGCTGGCGGAAGTTCCGAGAGNACTTACACCTTGGCGGTCAACAATACC 2976  
641 snLeuValAsnAlaGlyAsnSerAlaSerGlyLeuValaThrSerGlyLys 657  
2977 GGCACAGAACCCGTA...AGCTCGATCAATTGACGAGTAGTGAAGGAA 3023  
658 GlyIleGlnValValaGlnAlaIleasnglyAlaThrThrGlnGlyGly 674  
3024 AGACACAACAAACGCTGTCGAAAACTTAATTCAACCCGCAAAAGCAAC 3073  
674 a.....PheValGlnGlyAsna 680  
3074 ACCTCATGCGCGCGCGGTGAGCTTACCAATC..... 3105  
680 rglLeuGlnAlaGlyAlaPheAsnTYrSerLeuAsnArgspSeraspGlu 696  
3106 .....ATCGCAAAAGACGGCGAGTTCCGCTGCAGCAATATCCGCTCA 3146  
697 SetTrpTYrLeuArgSerGlnAsnAlaTYrArgAlaGluValProLeuTY 713

---

3147 AGACACAAGAGCTTTCGCAAACTC.....GCCA 3175  
713 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
ralasMetLeuThrGlnAlaMetAspTYrArgAlaValAlaGlys 730  
3176 AGCGACAGAACCCAAAACAGCGCGGAGAAAAGCAACGCCAAAGCCTTAC 3225  
730 eArgSerHisGlnThrGlyValaInglyGluAsn..AsnSerValArg 745  
3226 GCGCTGATTGCGCGCGCGCGATGCCGCCGCAAAAGACAGAAAGCTTGC 3275  
746 LeuSerIleGlnGlyGlyHisLeuGlnHisAspAsnAsnGlyGlyIleAl 762  
3276 CGAACCGCGCGCGCGCGAGGCGGGGAAAATGTCCGCTATTTCAGCGCG 3325  
762 aArgGlyAlaThrProGlnSerSerGlySertylGlyPheValArgLeug 779  
3326 AGAAGAGAAAAAACGGTGCAG..... 3348  
779 lngLYAspLeuMetArgThrGluValAlaGlyMetSerValThrAlaGly 795  
3349 .....GGCGATAAGACAGCGCCNTTGGCGAAGACGCCGGAAC 3386  
796 ValTYrGlyAlaAlaGlyHisSerSerValAspValLysAspAspArg 812  
3387 GGAACCGCGCGCGGTATACACCGCTTCCCGCGCGCGCGCGCGCGCG 3436  
812 Yser.....ArgAlaGlyThrValArga 820  
3437 GGGAATTTCGCGAACCGGACGCCCAACCGCAACTCAACCCCAACCGCAG 3486  
820 spASP..... 821  
3487 CGCGACGTATNAGCGCTTATGCCAATAGCGGTTGATGAATTTTCGCG 3536  
822 .....AlaGlyLys.....LeuGlyGly 827  
3537 CACGCTACACAGCGTTTTGCGCGGTACAGAGAGCAATGAGCGCGTGTGG 3586  
827 TYrLeuAsnLeuValHisThr..... 834  
3587 CGGAAGACCGCGCGACGCGNTTGGACAAACGNCATC...CGGNACACG 3633  
835 .....SerSerGlyLeuThrPlasplileValAlaGlnGlyThr 847  
3634 AAACACTACGCTTGCAAGATTTCCGCGCTTACCGCGCAACAAACGCACT 3683  
848 ArgHis.....SerMetLysAlaSerSerAspAsnAspPh 860  
3684 GCGCCAAATCGTATGCAAGAAAACCTCGGACAGGCGCGGTGGCGATCC 3733  
860 eArgAlaArgGlyThrPlgLyrrp...LeuGlySerLeuGlnThrGlyLeup 876  
3734 TGTTTTGGCACACCGGACGGAAC..... 3759  
876 ropHeSer.....IleThrAspAsnLeuMetLeuGlnProGlnLeuGln 890  
3760 .....ANCOTGACAGAGGAGATCGGCAACCTCGGCAGC 3791  
891 TYrThrTrpGlnGlyLeuSerLeuAspAspGlyLysAspAsnAlaGlyTy 907  
3792 GCTTGGCCACGCGCGCTTTTCGGGCAATAGCGCATCGCAGGAGTTGACA 3841  
907 rValLysPheGlyHisGlySerAlaGlnHisValArgAlaGlyPheArgL 924  
3842 TCggc.....ATCAGACAGCGCGCGGCTTTAGCACG...GGC 3876  
924 euGlySerHisAsnAspMetThrPheGlyGlnGlyGlyThrSerSerArgAla 940  
3877 ANTCTTTCAGACGCGCATCGGAGGCAAAATCCGCGCGCGGTGTCATTA 3926  
941 ProLeuArgAspSerAlaLysHisSerValSerGlnLeuProValAsnTr 957

```

3927 CGGCATTCAGGCGACGATACCGCGCGCTTCGGCGGATTCGCATCGAAC 3976
      : : : : :
957 pTPrValGlnPro..... 961
3977 CGTACATCCGCGCAACGGCATTTCTCCAAAAAGCGATTAACCGCTAC 4026
      : : : : :
962 .....SerValIleArgThrPheSerSerArgGlyAspMetArg... 974
      : : : : :
4027 GMAACGTCAATATCCGACCCCGGCTTTCGCTTCAACGATACCGNGC 4076
      : : : : :
975 .....ValGlyThrSerThrAlaGly..... 981
4077 GGGCATTAAGCAGATATTCATCAACCGCGCGCAACATNCCATCA 4126
      : : : : :
982 .....SerGlyMetThrPheSerProSerGlnAspGlyThr.... 993
4127 CNCCTATTNAGCGTGTCTATACGATCCGCTTCGGCGCAAGTCCGA 4176
      : : : : :
994 .....SerLeuAspLeuGlnAlaGlyLeuGlnAlaArgValArg 1006
4177 ACACGCGTCAATACCGCGCATTTGGCTCAGGATTTCCGCAAAACCGCAG 4226
      : : : : :
1007 GlnAsnIleThrLeuGlyValAla..... 1015
4227 TCGGGAATGGGGCGTAACGCCGAATCAAGTTTCACGCTGCATGCC 4276
      : : : : :
1016 .....GlyTyrAlaHisSerValS 1022

```

seq\_name: SwissProt\_40:PRT\_BORBR

seq\_documentation\_block:

ID PRT\_BORBR STANDARD; PRT; 911 AA.

AC 003035;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Pertactin precursor (outer membrane protein P.68) (P.94).

PRN.

OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).

OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;

CC Bordetella.

OX NCBI\_Taxid=518;

RN [1]

RP SEQUENCE FROM N.A. AND SEQUENCE OF 35-44.

RC STRAIN-ISOLATE CN7531;

RX MEDLINE=92407514; PubMed=1527510;

LA J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;

"Cloning, nucleotide sequence and heterologous expression of the

protective outer-membrane protein P.68 pertactin from Bordetella

bronchiseptica.";

RT J. Gen. Microbiol. 138:1697-1705(1992).

RL -1- FUNCTION: AGGLUTININ GEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS

MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN

BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.

CC -1- SUBUNIT: MONOMER.

CC -1- SUBCELLULAR LOCATION: Outer membrane.

CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.

CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW Mg(2+)

CC CONCENTRATIONS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation-

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X54815; CAA38584.1; -.

```

DR EMBL: A19180; CAA01453.1; -
DR PIR: A47675; A47675.
DR InterPro: IPR003992; pertactin.
DR InterPro: IPR003991; pertactin_related.
DR PRINTS: PRO1482; PERTACTIN.
DR PRINTS: PRO1484; PERTACTIN_FAMILY.
KW Outer membrane; signal; virulence; Repeat.
FT SIGNAL 1 34
FT CHAIN 35 911 P.94.
FT CHAIN 35 912 PERTACTIN (P.68).
FT PROPEP 713 911 POTENTIAL.
FT SITE 260 262 CELL ATTACHMENT SITE (INVOLVED IN
ADHESION TO VARIOUS EUKARYOTIC CELL
LINES).
FT SITE 701 703 CELL ATTACHMENT SITE (POTENTIAL).
FT DOMAIN 266 275 3 x 5 AA TANDEN REPEATS OF G-G-A-V-P.
FT REPEAT 266 270 1.
FT REPEAT 271 275 2.
FT REPEAT 276 280 3 (APPROXIMATE).
FT DOMAIN 570 601 7 x 3 AA REPEATS OF P-Q-P.
SQ SEQUENCE 911 AA; 93995 MM; 3078DF6C2D987A1 CRC64;

```

alignment\_scores: Quality: 224.00 Length: 1110  
Ratio: 0.475 Gaps: 51  
Percent Similarity: 42.523 Percent Identity: 20.000

alignment\_block: US-09-303-518d-651 x PRT\_BORBR

Align seg 1/1 to: PRT\_BORBR from: 1 to: 911

```

1279 GGGCGGGGGTTCATATCATGTAAGACAGTACGTTACTTGGAAAGTAA 1328
      |||||
58 GlyIleGlyValArgThrAlaThrGlyThrIle.....LysValSe 72
1329 CGGCGTGGCAAGACCGCGCTGCAAAATCGGCAAAAGCAGCGTCAGC 1378
      : : : : :
72 rGly.....A 74
1379 TTCAGCCAAAGG.....GAAACCAAGCGTCG...ATCAGCGTG 1416
      : : : : :
74 rGlnAlaGlnGlyValLeuLeuGlnAsnProIleAlaGluLeuArgPhe 90
1417 GCGGACGGTACAGTCAATTTGGATCAGCGCAGACGATTAAGCAAAAA 1466
      : : : : :
91 GlnAsnGlySerValThrSerSerGlyGlnLeuPheAspGluValArg 107
1467 ACAAGCCTTAGTGAAATCGGCTTGNTCAGCGCAGGGGTAGCGTCAC 1516
      : : : : :
107 gArgPheLeu.....GlyThrValThrV 115
1517 TGAATGGCGAATATCAAGTTCACACCGCAAAACTATTCGCGCTTCGC 1566
      : : : : :
115 allysAla..... 117
1567 GGGCGAGCTTGGATTAACGGGCAATTCGTTGCTCCAGCGTATCA 1616
      : : : : :
118 ...GlyLysLeuValAlaAspHisAlaThrLeuAlaAsnValSerAsp 133
1617 AAATACCGATGAAGGGCGATGTCGATGTCATTAATGCAACAAACAT 1666
      : : : : :
133 rArgAspAspAspGlyLeuAlaLeuThrValAlaGlyGluGlnAlaGln 150
1667 CCACCGTT.....ACCATTAACAGGATGAATGAATATTAACAACCG 1707
      : : : : :
150 lAspThrIleAlaAspSerThrLeuGlnGlyValAlaArgValGlu 166
1708 AGTGTAAGATATCAATAGACTTAATTAACGCAAAAGAAATGCTACAA 1757
      : : : : :
167 ArgGlyAlaAsnValThr..... 172

```

```
1758 CGGTGGTTGGCGAGAAAGATACGACCAAAACGACGGCGCTCAAC 1807
      ::::: ||| ::::: ||| ||:::
173 .....ValGlnArgSerThrIleValAspGlyGlyLeuHst 185
      ::::: ||| ::::: ||| ||:::
1808 TT.....GTTACCAAGCCCGCGCAGAGAACCGACCCGCTG 1845
      ::::: ||| ::::: ||| ||:::
185 IegLYThrLeuGlnProLeuGlnProGluAspLeuProProSerArgVal 201
      ::::: ||| ::::: ||| ||:::
1846 CTTTCGGCGGAGCAAAATTTAAAGGCAACATCCAGCAACCAACGCGCA 1895
      ::::: ||| ::::: ||| ||:::
202 ValLeuGlyAspThrSerValThrAla.....ValProAlaSerGly.. 215
      ::::: ||| ::::: ||| ||:::
1896 ACTGTTTTCAGCGCGCAGACCGCACCGCCTACATCATTTAGCA 1945
      ::::: ||| ::::: ||| ||:::
216 .....AlaProAlaAlaValSerValPheGlyAla 225
      ::::: ||| ::::: ||| ||:::
1946 GCGGGTGTCAAAATGGAAGGTATCCACAGAGAAATCGTGTGGAC 1995
      ::::: ||| ::::: ||| ||:::
225 IaAsnGluLeuThrValAspGly..... 232
      ::::: ||| ::::: ||| ||:::
1996 AACGACTGATCCMACCGCACGTTTAAAGCGAAATTCATATTCAAGG 2045
      ::::: ||| ::::: ||| ||:::
233 .....GlyHisIleThrGly 237
      ::::: ||| ::::: ||| ||:::
2046 CGGCGAGCGGTGATTTCCCGCATGTTGCCAAAGTGAGGCGATTGNC 2095
      ::::: ||| ::::: ||| ||:::
237 YGlyArgAlaAla.....GlyValAla..... 244
      ::::: ||| ::::: ||| ||:::
2096 ATTTAGCAATCAGCCCAAGCAGTTTGTGTGCGACCCCATCAAAAGC 2145
      ::::: ||| ::::: ||| ||:::
245 .....AlaMetAspGlyAlaIleValHisLeuGln 254
      ::::: ||| ::::: ||| ||:::
2146 CATACATCTGTACAGCTTTCGAG..... 2169
      ::::: ||| ::::: ||| ||:::
255 ArgAlaThrIleArgArgGlyAspAlaProAlaGlyAlaValProGly 271
      ::::: ||| ::::: ||| ||:::
2170 .....TGAGCNGGTC 2179
      ::::: ||| ::::: ||| ||:::
271 YGlyAlaValProGlyGlyPheGlyProLeuLeuAspGlyTrpTyrGlyVal 288
      ::::: ||| ::::: ||| ||:::
2180 TGACAAATTTGTGCAANAANCAATTACCGACGATTAAGTGATGCTTCA 2229
      ::::: ||| ::::: ||| ||:::
288 al.....AspValSerAspSerThrVal..... 295
      ::::: ||| ::::: ||| ||:::
2230 TTGACTAGACNGACNTNAGCGGCANTGTNAGCTNNCCNATNAGCNTNN 2279
      ::::: ||| ::::: ||| ||:::
296 .....AspLeuAla 298
      ::::: ||| ::::: ||| ||:::
2280 TTNAANCTCNCNGGCGNTGCNCACTNAANGCAATCTTAGTCGCAATG 2329
      ::::: ||| ::::: ||| ||:::
298 agInserIleValGluAlaProGlnLeuGlyAlaIleIleArgAlaGlyAla 315
      ::::: ||| ::::: ||| ||:::
2330 GCGATACAGCTTATACAGTCCGACAC.....GCCACCCCAAAAGCGC 2373
      ::::: ||| ::::: ||| ||:::
315 rgGlyAlaArgValThrValSerGlyGlySerLeuSerAlaProHisGly 331
      ::::: ||| ::::: ||| ||:::
2374 AACCTTAGCCTCTGGGCAATGCCCAAGCAACATTTAAT..... 2412
      ::::: ||| ::::: ||| ||:::
332 AsnValIleGluThrGlyGlyAlaArgArgPheProProProAlaSe 348
      ::::: ||| ::::: ||| ||:::
2413 .....CAAGCCACATTTAAAGCGCAACNCAATCGANTTGGGCAATGCT. 2454
      ::::: ||| ::::: ||| ||:::
348 rProLeuSerIleThrLeuGlnIleValAlaArgAlaGlnIleArgAla 365
      ::::: ||| ::::: ||| ||:::
2455 .....TCATTTAATCTAAGCAACAC 2475
      ::::: ||| ::::: ||| ||:::
365 euLeuTyrArgValLeuProGluProValIlySLeuThrLeuAlaGlyGly 381
      ::::: ||| ::::: ||| ||:::
2476 GCCGCAAAAAGCGAGTGTGAGCTTTCGCAACAAGCTAAGGCAAAAGT 2525
      ::::: ||| ::::: ||| ||:::
382 AlaGlnGlyGlnGlyAspIleValAlaThrGluLeuProProIle..... 396
      ::::: ||| ::::: ||| ||:::
2526 AAGCATTCGCGACTCAACGGC.....AATGTCTCCCTAGCGATAGG 2569
      ::::: ||| ::::: ||| ||:::
      ::::: ||| ::::: ||| ||:::
397 ....ProGlyAlaSerSerGlyProLeuAspValAlaLeuAlaSerGlnAla 412
      ::::: ||| ::::: ||| ||:::
2570 CAGTATTCAATTTTGAACAGCCGCTTTACCGGCAACTACGCGGCAC 2619
      ::::: ||| ::::: ||| ||:::
412 Ia.....ArgThrProGly.....AlaThrArg 419
      ::::: ||| ::::: ||| ||:::
2620 AAGGANAACGATTTACCTTAAAGACAGGATTGACGCTGCCGACG 2669
      ::::: ||| ::::: ||| ||:::
420 AlaValAspSerLeuSerIleAspAsnAlaThrTrpValMetThrAspAs 436
      ::::: ||| ::::: ||| ||:::
2670 CAGCAATTTAGCAATTTAAACCTTGACAGCCGACCATTAACCTCAAT 2719
      ::::: ||| ::::: ||| ||:::
436 nSerAsnValGlyAlaLeuArgLeu..... 444
      ::::: ||| ::::: ||| ||:::
2720 CCGCTATTCGCCACGATGCTGACAGGCGCGCAACCGGACAGTGTACAG 2769
      ::::: ||| ::::: ||| ||:::
445 .....AlaSerAspGlySerVal..... 450
      ::::: ||| ::::: ||| ||:::
2770 ACGCCGCGCGCGCTTCGCGCGCTTCCTTATTCGCTTACACCGGCAC 2819
      ::::: ||| ::::: ||| ||:::
451 .....AspPheGlnGlnProAla 456
      ::::: ||| ::::: ||| ||:::
2820 TTGCGTAGAATCCCGTTCACACAGCTGACGCTAAACGGCAATGACN 2869
      ::::: ||| ::::: ||| ||:::
456 agIVala..GlyArgPheIlyScyLeuMetValasp..ThrLeuAlaG 471
      ::::: ||| ::::: ||| ||:::
2870 GTCAGGACATTCGCTTATGTGCGAACTCTTCGCGCTACGACGAC 2919
      ::::: ||| ::::: ||| ||:::
471 LysSerGlyLeuPheArgMetAsnValPheAlaAspLeuGlyLeuSerAsp 487
      ::::: ||| ::::: ||| ||:::
2920 AAATGGAAGCTGGGGAAGTCCGAAGGACCTTACACCTTCGCGGTCA 2969
      ::::: ||| ::::: ||| ||:::
488 LysLeuValValMetLysAspAlaSerGlyGlnHisArgLeuLeuValAs 504
      ::::: ||| ::::: ||| ||:::
2970 CAATACCGGCAACGACCCGTAAGCCTCGATCAATTAGCGTAGTGAG 3019
      ::::: ||| ::::: ||| ||:::
504 gAsnSerGlySerGluProAlaSerGlyAsnThrMetLeuLeuValGln. 520
      ::::: ||| ::::: ||| ||:::
3020 GGAAGACAAACACCGCTCCGAAACCTTATTTACACCTCGCAAAAC 3069
      ::::: ||| ::::: ||| ||:::
521 .....ThrProArgGlySerAlaAlaThrPheThrLeuAlaAsn 533
      ::::: ||| ::::: ||| ||:::
3070 GAA.....CAGCTGATGCGCGGCGGTGCTTACCACTCATCGCA 3113
      ::::: ||| ::::: ||| ||:::
534 LysAspGlyIlyValAspIleGlyThrTyrArgTyrArgLeuAlaIAs 550
      ::::: ||| ::::: ||| ||:::
3114 A..GACGGGAGTTCGCGCTGCATATCGGTCAAAAGAACAGAGCTTT 3160
      ::::: ||| ::::: ||| ||:::
550 ngIyAsnGlyGlnTrpSerLeuVal..... 558
      ::::: ||| ::::: ||| ||:::
3161 CCGACAAACTCGGCAAGGACGAAAGCCAAAAAACAGCGGAAAAAGACAC 3210
      ::::: ||| ::::: ||| ||:::
558 ..... 558
      ::::: ||| ::::: ||| ||:::
3211 GCGCAAGCCTTGACGGGTGATTGGCGCGGCGGATGCCGCGAA 3260
      ::::: ||| ::::: ||| ||:::
559 .....GlyAlaIlySalProProA 565
      ::::: ||| ::::: ||| ||:::
3261 GACAGAAAGCGTTGCCGACCGCGCGCGGACGAGCGGAAATGTG 3310
      ::::: ||| ::::: ||| ||:::
565 IapTrpLysProAlaPro...GlnProGlyProGlnProGly..... 577
      ::::: ||| ::::: ||| ||:::
3311 GCATTATGCAAGGCGGAGAGAAAGAAAAACGGGTGCAGCGGTAAGAC 3360
      ::::: ||| ::::: ||| ||:::
577 ..... 577
      ::::: ||| ::::: ||| ||:::
3361 AGCGCNTTGGCAAAACAGCGCGAAGCGAAACCGCGGNTACACCGC 3410
      ::::: ||| ::::: ||| ||:::
578 .....Prog 579
      ::::: ||| ::::: ||| ||:::
3411 CTTTCCCGCGCGCGCGCGCGCGGATTTTGCAGCAACGCGACGCGC 3460
      ::::: ||| ::::: ||| ||:::
```

```

579 In.ProfProGlnProGlnProPro.GlnProProGlnInProGlnProG 595
3461 AA...CCGCAACCTCAACCCCAACGCGACGCGCATGATNAGCCGTTAT 3507
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
595 LuAlaLeuGlyArgGlyHisSerLeuTyraLaseTyrGlyTyrSerLys 610
3508 GCCAATAGCGGGTTGAGTAATTTCCGCGCACGCTCAACAGCGTT... 3552
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
611 AlaAsnAlaAlaValAsnThrGlyValGlyLeuAlaSerThrLeuTr 627
3553 .TTCCGCCGACAGAGAGATTTGACCGCGCTTTTCCGAGACCGCCGCA 3601
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
627 pTyraLagLiuSerAsnAlaLeuSerLysArgLeuGlyLiuLeuArgLeu 644
3602 AC.....GCGTTTGACAAACGCGATCCGCGNACCAACAC 3639
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
644 snProAsnAlaGlyValaTrrpGlyArgGlyPheAlaGlnArgGlnGln 660
3640 TACCGTTGCAAGATTTCGCGCCTACCGCAACAACGACCTCGGCA 3689
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
661 LeuAspAsnArgAlaGlyArgPheAsnArgGlnLysValAlaGlyPheG 677
3690 AATCGGTATGCAAAAAACCTC.....GCGACGCGCGCGCTCG 3727
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
677 uLeuGlyAlaAspHisAlaValAlaValAlaGlyArgTrrpHisLeuG 694
3728 GCATCGCTGTTTGGCACACGCGCAAGAAACATTGACAGACGGCATC 3777
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
694 LysGlyLeuAlaGlyTrrpHisArgGlyAspArgGlyPheThrGlyAspG 710
3778 GCGCACTGCGCAAGCGCTTCCGCGCGCGCTTTCGCGCAATACGCGAT 3827
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
711 GlyGlyHisThrAspSerValHis.....ValGlyGlyTyrAlaThr 724
3828 CCGCAGGTTGACATGCGCATGACAGACGCGCGGTTTACACAGCGCA 3877
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
724 rTyrrileAlaAsnSerGlyPheTyrrleuAspAlaThrLeuArgAlaSe 741
3878 NCTN.....TCAGACGCG.....ATCGGA 3897
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
741 rGleuGlnAsnAspPheLysValAlaGlySerAspGlyTrrAlaValLys 757
3898 GCGCAAAATCCGCGCGCGCTGTCATTAAGCGATTCAGCAGATACCG 3947
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
758 GlyLysTyrArgThr.....HisGlyValGlyAlaSerLeuGln 770
3948 CCGCGCGT.....TTCCGCGGATTCGCGCATGCAACCGTTACA 3982
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
770 uAlaGlyArgArgPheAlaHisAlaAspGlyTrrPheLeuGlnProGlnA 787
3983 TCGGCGCAGACGCGCTATTGTCGCAAAAAGCGATTAACCGCTACGAAA 4032
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
787 lagLiuLeuAlaValPheArgValGlyGlyGlySerTrrArgAlaAlaAs 803
4033 ...GTCAATATCGCACCCCGCGCTTGGCTTCAACCGTACGCGGCGG 4079
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
804 GlyLeuArgValArgAspGlnGlyGlySerSerValLeuGlyArgLeuG 820
4080 CATTAGGAGATTAATTCATTAACCGCGCGCAACATNTCCATCACAC 4129
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
820 yLeuGlnValGlyLysArgGlyGlyLeuAlaGlyArgGlnValGln 837
4130 CTTAATTNAGCGCTGCTTACGAGTGGCTTCCGCGCAAGTCCGAGACA 4179
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
837 rGlyTrrileLysAlaSerValLeuGlnGlnPheAspGlyAlaGlyThrVal 853
4180 CCGGTCAATACCGCGCTATTGGCTCAGATTTCCGCAAAAACCGCAGTGC 4229
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
854 ArgThrAsnGlyLysAlaHisArgThrGlnLeuArgGlyThrArg...Al 869
4230 GGAATGCGCGCTAAACGCGCAAAATCAAGTTTCAACGCTGCTCCNTCCACG 4279
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
869 agLiuLeuGly.....LeuGlyMetAlaAla 877

```

```

4280 CTGCGCGCGCCAAAGNCCGCAACTGGAACG.....CAACACAGCGCG 4323
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
877 lAlaLeuGlyArgGlyHisSerLeuTyraLaseTyrGlyTyrSerLys 893
4324 GGCATCAATTAAGCTACCGCTCG 4347
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
894 GlyProLysLeuAlaMetProTrrp 901

seq_name: SwissProt_40:HLVA_PROM1

seq_documentation_block:
ID HLVA_PROM1 STANDARD: PRT: 1577 AA.
AC P16466;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Hemolysin precursor.
GN hPMA.
OS Proteus mirabilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Proteus.
OX NCBI_TaxID=584;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-43.
RC STRAIN-ISOLATE 477-12;
RX MEDLINE=90170827; PubMed=2407716;
RA Uphoff T.S., Welch R.A.;
RT Nucleotide sequencing of the Proteus mirabilis calcium-independent
RT hemolysin genes (hPma and hPmb) reveals sequence similarity with the
RT Serratia marcescens hemolysin genes (shla and shlb).";
RL J. Bacteriol. 172:1206-1216(1990).
CC -! FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
CC CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
CC DEFINED.
CC -! FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
CC ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. hPMA
CC REQUIRES hPMB FUNCTION.
CC -! SUBCELLULAR LOCATION: Outer membrane.
CC -! MISCELLANEOUS: THE CONSERVED AMPHIPATHIC DOMAINS IN SHLA AND hPMA
CC MAY BE RESPONSIBLE FOR PORE FORMATION.
CC -! SIMILARITY: TO S. MARCESCENS HEMOLYSIN (SHLA).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M30186; AAA25657.1; -.
DR PIR: A35140; A35140.
KW Hemolysins; Toxin; Outer membrane; Signal.
FT SIGNAL 1 29
FT CHAIN 30 1577 HEMOLYSIN.
SQ SEQUENCE 1577 AA; 165869 MW; 175975E0C924B2D9 CRC64;

```

alignment\_scores:
 Quality: 217.00 Length: 1682
 Ratio: 0.290 Gaps: 82
 Percent Similarity: 44.530 Percent Identity: 18.906

alignment\_block:
 US-09-303-518D-651 x HLVA\_PROM1

Align seg 1/1 to: HLVA\_PROM1 from: 1 to: 1577

```

40 AAAGCCCGCAAAACCGCGCGATCCGCTTCGCGCTGCTACTTACGCAAT 89
   ||| .....||| ||| ||| ||| ||| ||| ||| ||| |||
7 LysLeuSerProSerGlyArgLeuAlaAlaSerLeuAlaIleIlePheVa 23

```





576 .....GlnGlyAlaPheValLysThrThrIng 585  
1304 ACAGTACCGTACTTGGAAAGTAAAGCGCGTGGCAAAAGACCGCTGTC 1333  
.....  
585 LysPValVal.....IleAspAlaLeuSerGluThrIleSer 598  
1354 AAAATCGGCAAA.....GGCAGCGTCGACGTTCAAGCCAAAGGGGAAA 1397  
||||| .....  
599 LysIleAspGluArgThrGlyThrAlaPheAsnIleThrLysSerSerI 615  
1398 CCA.....GGCTCGATCAGCGTGGGCG 1420  
.....  
615 sLysnGlnIuThrAsnLysGlnThrSerThrGlySerGluLeuIleSerA 632  
1421 ACGGT.....ACAGTCATTTGGATCGAGCGGACAGATAAAGGCAA 1464  
||| .....  
632 spAlaGlnLeuThrValValSerGlyAsnAspValAsnValIleGlySer 648  
1465 AAA.....CAAGCCTTAGTGAATCGGCTTGNTCAGCGGACGGGTACG 1511  
.....  
649 LeuIleLysSerAlaAspLysLeuGlyIleHisSer..LeuGlyAspI 664  
1512 GCACTGAATGCCGTAATCAG.....TTCAACCCGACAACTCT 1552  
.....  
664 eAsnValLysSerAlaGlnGlnValThrLysIleAspAspGluThrS 681  
1553 ATTTCGGCTTCGGCGGACGTTGGATTAAACGGG.....CATTCG 1596  
.....  
681 eRLeuAlaIleThrGlyHisAlaLysGluValAspLysGlnLysSer 697  
1597 CTTTCGTTCCACCGTATTCAAAATACGATGAAGGGCGGATGATGNCNA 1646  
.....  
698 AlaGlyPheHisIleThrHisThrAsnLys..... 708  
1647 TCATATGCCSACACACATCCACGTTACCTTACAGGAATGAATA 1696  
.....  
709 .....AsnThrSerThrGluThrGlnIleAsnSerThrI 721  
1697 TTACACAAACCGAGTGAAGTATCATATACACTTAATTAACGCAAA 1746  
||| .....  
721 Le.....SerGlyAlaAsnValAsp..LeuGlnAlaAsnLysAsp 733  
1747 ATTTCCTACACGCTTGGTTGGCGAAGAAAGTACGACCAAAACACAGG 1796  
.....  
734 ValThrPheAlaGly.....SerAspLeuLysThrThrAlaGlyAsnAl 748  
1797 GGGGCTC.....AACCTGTTTAC.....CAGCCG 1822  
.....  
748 AserIleThrGlyAspAsnValAlaPheValSerThrGluAsnLysG 765  
1823 CGCAGAAAGACCGCACCCNGCTGCTTCCGGCGGAACAATTTAAAGGC 1872  
.....  
765 IuThrAspAsnThrAspThrThrIleSerGlyPheSerThrGly 781  
1873 AACATACCGCAAAACAAACGCAACTGTTTTCACGGGACGACGACAC 1922  
.....  
782 GlyValAspLysValGlySerLysAlaAspPheGln..... 793  
1923 GACGCGCTCAATCATATTAGAAAGCGGCTGTCAAAAATGAAGGTATCC 1972  
.....  
794 .TyrAspLysGlnHisThrGlnThrGluValThrLysAsnArgGly...S 809  
1973 CACAAAGAGAAATCGTGTGGACAAACGACTGATCACCACGCTTTAA 2022  
||| .....  
809 eRGlInThrGluValAlaGlyAsp.....LeuThrIleThrAlaAsn 822  
2023 GCGGAATTTCCATTTACAGGCGGCGAGCGGATTTCCGGAATCT 2072  
.....  
823 LysAspLeuLeuHisGluValAserHisHisValGluGlyArgGly 839  
2073 TGCCAAAGTGAAGCGATTTGACATTTGACATCAGCCCAAGAGATT 2122  
.....  
839 nGluSerGlyGluAsnIleGlnHisLeu..... 848  
2123 TTGCTGCGACCGCATCAAGCCATACATCTGTACAGCTTCG..... 2166  
||| .....  
849 .....AlaValAsnAspSerGluThrSerLysThrAspSerLeuAsn 862  
2167 .....GACTGACNGCTGTGACAAATG 2189  
||| .....  
863 ValGlyIleAspValGlyValAsnLeuAspTyrSerGlyValThrLysPr 879  
2190 TTGCGAANAANACATTACGACGATTAAGTATGCTGTCATGACAAACA 2239  
||| .....  
879 oValLysLysAlaIleGluAspGly.....ValAsnThrThrLysP 893  
2240 CNGACNTNAGCGCANTGTNAGCNTNCCNATNACGNTNNTNAAANTC 2289  
.....  
893 roGlyAsnAsnThrAspLeuThrLysLysValThrAlaArgAspAlaIle 909  
2290 NCNGGCGCTGCNNCCTNAN..... 2310  
.....  
910 AlaAsnLeuAlaAsnLeuSerAsnLeuGluThrProAsnValGlyValG 926  
2311 .....GGCAATCTTAGTGAATGCGGATACAGT 2340  
.....  
926 uValGlyIleLysGlyGlyLysSerGlnGlnSerGlnThrAspSerGlnA 943  
2341 ..TATACAGTCACGCAACAGCC..... 2361  
.....  
943 lValSerThrSerIleAsnAlaGlyLysIleAspIleAspSerAsnAsn 959  
2362 .....ACCCAAACGCGCAACT 2378  
.....  
960 LysLeuHisAspGlnGlyThrHisTyrGlnSerThrGlnGlnGly...I 975  
2379 TAGCTCTGCGCAATGCC.....C 2398  
.....  
975 eSerLeuThrAlaAsnThrHisThrSerGluAlaThrLeuAspLysHisG 992  
2399 AACCAATTTAAATCAAGCCATTAACGCGCAACMCATCGCTGGGCG 2448  
||| .....  
992 IuThrThrPheHisGluThrLysGlyGlyGlyIleGlnValSerThr 1008  
2449 AATGCTCATTTAATCTAACCAACAGCGCGCAAAACGCGAGTGTGAC 2498  
.....  
1009 LysThrGlySerAspIleThr.....ValAlaIleLysGlyGlnGly 1023  
2499 GCTTTCGACACGCT.....AAGCAAAAGTAAAGCATTCG 2536  
.....  
1023 nThrThrAspAsnAlaLeuMetGluThrLysAlaLysGlySerGlnPheT 1040  
2537 CACTCAACGCGCAATGCTCCCTGCGGATGAAGCAGATTCATTTGAA 2586  
.....  
1040 hrSerAsnGlyAspIleSerIleAsnValGlyGluAsnAlaHisTyrGlu 1056  
2587 AACACCGCTTTACCGGACAACCTCAGCGGACGAGCAAGANACACATTACA 2636  
.....  
1057 GlyAlaGlnPheAspAlaGln..... 1063  
2637 CTTAAAGACAGCGAATGACGCTGCCGTGACGACGAGAAATTAGCAATT 2686  
.....  
1064 .....LysGlyLysThrValIleAsnAlaGly.....GlyAspL 1075  
2687 TAAACCTTGACAAAGCGCACCATTTACACTCAATTCGCGTATCGCAGAT 2736  
||| .....  
1075 eutnRLeuAlaGlnAlaThr..... 1081  
2737 GCTCAGGCGCGCAACCGGACGNGTCTCAGACAGCGCGCGCGCTTC 2786  
.....  
1081 ..... 1081  
2787 GCGCGCTTCCTATTATCCGTTACACGCCCAACTGCGTAGAATCCCGTT 2836  
||| .....  
1082 .....AspThrHisSerGluSerGlnS 1089

```

2837 TCACACGGCTGACGGTAACGGCAATTAACNGTCAGAACATTCCCG 2886
      |||
1089 erasn.....ValasnGlySerAlaasn..... 1096
2887 TTATATCGGAACCTCTTGGCTAACGACGACGCAAAATTAACCTGGCGGA 2936
1097 .....LeuLysValGlyThr 1101
2937 AAGTTCGGAAGNACTTACACCTTGGCGGTCAACATACCGCAACGAGAC 2986
      |||
1101 rthrProGlySerLysAspTyrGlyGlyLysPheAsnAlaGlyThrThrH 1118
2987 CCCTAAGCCTCGATCAATTGACG.....GTACTGGAAGGAAA 3024
      |||
1118 LshIsSerLysGlyGlnThrThrAlaLysValGlyThrThrLeuThrGlySer 1134
3025 GACAAACAAACCGCTGCGCAAAACCTTATTCACCTCCCAACAGCAACA 3074
      |||
1135 GlnGlyLLeuLysValAsnAlaGlyHisAsnLeuThrLeuGlnGlyThrH 1151
3075 CGTCGATGCGCGGCGCTGGCTTACCACTCATCCGCAAGAGCGGAGCT 3124
      |||
1151 sleuSerSer.....GluGlnAspLLeaAla 1160
3125 TCCGCTCGATTAATCCGCTC..... 3144
      |||
1160 euAsnAlaThrAsnLysValAspLeuGlnSerAlaSerSerGlnHisThr 1176
3145 ...AAGAACAAGAGCTTCCGCAACACTCGCGCAAGCGCAAGCAAA 3191
      |||
1177 GlnLysGlyLysAsnAsnLeuSerGlyLysValGlnAlaGlyPheGlyLysLys 1193
3192 ACAGGCGGAAAAAGACAAACGCAAGCCTTGACGCGCTG..... 3231
      |||
1193 smethr.....AspAspAlaSerSerValAsnGlyLeuGlySerAlaG 1208
3232 ..ATTGCGCGCGCGCGCGATGCCCGCAAAACAGCAAAACCGTTGCCGA 3279
      |||
1208 LnpheAlaLLeuGlyLysGlnAspGlyLysSerValSerArgLysGly 1224
3280 CCGGCGCGCGCGCGAGCG.....GGGGAATAATGTCGGCAT 3314
      |||
1225 ThrLLeaAsnSerGlyAsnLeuThrLLeaGlnLysSerValHisG 1241
3315 TATGCAAGCGGAGGAAGAAAAACGGTGCACGCGGTAAGACAGCG 3364
      |||
1241 uGlnGlyAlaGlnValAsnSerLysAspThrGlnLeuThrSerGlnSerG 1258
3365 CMTTGGCGCAACAGCGCGCAAGCGGAACCGCGCGGNTACACACGCGCTTC 3414
      |||
1258 LysAspLLeuLLeuThrSerAlaGlnSerThrAspTyrLysAsnAsnTrp 1274
3415 CCGCGCGCGCGCGCGCGCGCGGATTTGCCGCAACCGCGCAACCGCCCAAC 3464
      |||
1275 GlyThrAspLLeuGlyPheAsnLysLysThrAsnAsnThrProLysG 1291
3465 GCACCTCAACCCCAACCGCACCGCGACCTG..... 3495
      |||
1291 uValThrGlnGlyLysProAlaThrSerLLeuHisAlaGlyLysL 1308
3496 .....ATNAGCCGTTATGCCAATAGCGCGT 3519
      |||
1308 euLeuValAsnValGlnAspGlnGlnLysThrSerHisGlnAsnAlaThr 1324
3520 TTGAGTGAAATTTCCGCAACGCTCAACAGCGTTTCCGCGCGTAAGAGAGA 3569
      |||
1325 LeuGlyThrGlyThrLeuThrLLeaSer.....AsnLysAs 1337
3570 ATTGACCGCGCTGTTGCCGAAGCGCGCAACCGCTTGGACAGACGN 3619
      |||
1337 PheuThrLeuSerGlyAlaAsnValThrAlaAspSerVal...ThrGlyA 1353
3620 GCATCGCGNACACCAACACTACGCTTGCAGAAATTTCCGCGCGCTACCGC 3669

```

```

1353 snValGlyGlySerLeuAsnLLeaSerGln..... 1363
      |||
3670 CAACAACCGGACCTGCGCCAAATCGGTATGCGAAGAAACCTCGGACGG 3719
      |||
1364 LysGlnSerAsp...ArgHisVal..... 1370
3720 GCGGCTGGCATCCGTTTTCGCAACCGGACCGCAAAACNCTTCGACG 3769
      |||
1371 ThrValGlyValAsnValGlyTyrAsnHisThrAsnAspProLysSer 1387
3770 ACGGCATCGCACTCGGACGCGCTTGGCCAGCGCGCTTTCGGGCA 3819
      |||
1387 erGlnValAsnLysThrAlaLys...AlaGlyLysSerLeuGlnLys 1402
3820 TACGCATCGGACGTTGCATCGCATCGACGCGGCGGCTTTAG 3869
      |||
1403 ThrLLeuAspThrLLeaSerGlyLysSerSerThrAspAlaI 1419
3870 CAGC.....GGCANTCTNTCAGACGCGATCGGAGGCAAAATCCGCC 3910
      |||
1419 eSerAspLysTyrAsnSerLeuSerSerThrLLeaAspLysThr... 1434
3911 GCCGCTGCTGCATTAACGCGAT.....CAGCACGATACCGC 3948
      |||
1435 .....GlyLLeSerAspGlnThrLysAlaLysLLeaAsp 1445
3949 GCCGCTTTCGCG.....GGATTCGCATCGACCGTACATCGCGCAAC 3992
      |||
1446 GlnGlyPheGlyLysValGlyAsnGlyLLeuLysAsnLLeaValThrGlyAl 1462
3993 GCGCTATTTCGTCGCAAAAGCGGATTTACCGTACGAAACGTCAT 4038
      |||
1462 aGlnGlyHisThrAlaAsnAlaAspLLeuLysValThrHisValAsp 1477

seq_name: SwissProt_40:ALDA_ECOLI
seq_documentation_block:
ID ALDA_ECOLI STANDARD; PRT; 1286 AA.
AC Q03155;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Adhesin alda-I precursor.
GN ALDA-I.
OS Escherichia coli.
OG Plasmid p186.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 50-56.
RC STRAIN=2787 (O126:H27);
RC MEDLINE=92326638; PubMed=1625582;
RA Benz I., Schmidt M.A.;
RT "ALDA-I, the adhesin involved in diffuse adherence of the
RT diarrhoeagenic Escherichia coli strain 2787 (O126:H27), is
RT synthesized via a precursor molecule.";
RL Mol. Microbiol. 6:1539-1546(1992).
CC -1- FUNCTION: THIS IS AN ADHESION PROTEIN NECESSARY FOR THE DIFFUSE
CC ADHERENCE OF DIARRHEA-CAUSING ENTEROPATHOGENIC ESCHERICHIA COLI
CC TO EPITHELIAL CELLS.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X65022; CAA46156.1; -

```

DR PIR: S28634; S28634.  
 KW Cell adhesion; Signal; Outer membrane; Plasmid.  
 FT SIGNAL 1 49 ADHESIN AIDA-I.  
 FT CHAIN 50 ?  
 FT PROPER ? 1286  
 SO SEQUENCE 1286 AA; 132271 MW; B2A0E72AC05FB34 CRC64;

alignment\_scores:  
 Quality: 214.50 Length: 1456  
 Ratio: 0.341 Gaps: 62  
 Percent Similarity: 43.201 Percent Identity: 18.475

alignment\_block:  
 US-09-303-518D-651 x AIDA\_ECOLI ..

Align seg 1/1 to: AIDA\_ECOLI from: 1 to: 1286

```

646 TCCGGCGCATGGTTAATGCGGCAATACATATG.....CAGGGTTG 689
|||||: : : : : : : : : : : : : : : : : : : : : : : : :
58 SerGlyLThrValSerSerGlyLThrGlnIleValIyrSerGlyYar 74
690 GGAATATGCGGTANTAGTTTGCGCGCATGTGCCCATGGCCAGC 739
|||||: : : : : : : : : : : : : : : : : : : : : : : : :
74 ACTATGCGCCCTATGCCGATTCGAGTGCAGCGCAGCGGCTTCGCA 789
: : : : : : : : : : : : : : : : : : : : : : : : : : :
91 sn...GlyGlyLThrAlaThrValAsnSerGlyLThrGlnIleValAsn 91
790 ATGTTATTATGACAAACAAACAATAATGCGTCTGCAAC..... 831
107 AsnValGlyThrSerGlyAlaThrIleSerThrIleValAsnSerGly 123
832 .....GGAGTTTACAACCGGCTACCCCTTAT 859
123 yllleGlnArgValSerSerGlyLysAlaSerAlaThrAsnLeuSerG 140
860 CCGGCGAGGAGAAACGGTTCCAGCTGATACGCAAGATTGTTTCAGAT 909
: : : : : : : : : : : : : : : : : : : : : : : : : : :
140 LysGlnAsnIleValAsnLeu..... 148
910 GACATTACAGAGCGCATACATACCTCTTTTGAACCGCGCAGTAA 959
148 ..... 148
960 CGGACATTTTCTCTTACA.....TCCACACACACGGTAGC.... 996
149 .GlyHisAlaSerAsnThrValIlePheSerGlyLysGlnIlePhe 165
997 .....GTTACGCTACAGAACCGCAAGAGTTCATCCAAAGCTT 1041
165 heserGlyLThrIlePheSerThrAsnIleSerSerGlyLysGln 181
1042 AAGGTACAGACGTCGACTGTTTGCAATCT..... 1074
182 ArgValSerSerGlyLysAlaSerAsnThrIleAsnSerSerGly 198
1075 .....TTGAATGAACATGATAAGACAGTTTACCGCGCAG 1111
198 yAlaGlnAsnIleLeuSerGlnIleValIleSerThrHisIleSers 215
1112 GGGGTGTTAATCAGTAC..... 1128
215 erGlyLysGlnIleSerAlaGlyAlaAsnAlaThrGlnIle 231
1129 .....CGTCCAGGTTAAACACGCT..... 1149
232 ValAsnSerGlyLysPheGlnArgValAsnSerGlyAlaValAlaThr 248
1150 .....GAAACCTTTTATCATGATTCAGTACGGCA 1177
248 yThrValLeuSerGlyLThrGlnAsnValSer.....Serg 261

```

```

1178 ACGGCAACTCATCTTATCAACACATCAACAGCGCGCGGCTTG 1227
: : : : : : : : : : : : : : : : : : : : : : : : :
261 LysIleSerAlaIleSerThrSerValIleAsnSerGlyValGlnThrVal 277
1228 TATTTGCAAGT.....GATTTACGGTCTCGGCTTAACAA 1265
278 ...PheAlaGlyAlaThrValThrAspThrThrValAsnSerGlyLys 293
1266 CGAAACGTGGCAAGCGCGGCGTTCATATCAAGCAAGACATACGCTTA 1315
: : : : : : : : : : : : : : : : : : : : : : : : :
293 GlnAsnIleSerSerGlyLysIle...ValSerGlu.....Thr 306
1316 CTTCGAAAGTAAACGCGGTGCAAC.....GACCGCTG 1350
: : : : : : : : : : : : : : : : : : : : : : : : :
306 hrValAsnValSerGlyLThrGlnAsnIleIyrSerGlyLysAlaLeu 322
1351 TCCAAATCGGCAAGGACGCTGCACGTTCAAGCCCAAGGGAA.... 1395
323 SerAlaAsnIleLysGlySerGlnIleValAsnSerGlnIleThrAla 339
1396 .....AACCAAGGCTCA 1408
339 eAsnThrLeuValSerAspGlyLThrGlnHisIleArgAsnGlyLys 356
1409 TCACGCTGGCGACGCTACATCTTTGATCAGCAGCAGACGATAA 1458
: : : : : : : : : : : : : : : : : : : : : : : : :
356 LeuIleSerGly.....ThrIleValAsnGlnSerGlyLThrValAsn 369
1459 GGCAAAACAGCCTTAGGAATATGCGCTGTGTCGCGGAGGGGTAC 1508
370 IleSerSerGlyLysIleValGlnIleSerThrIleIleAsnSerGlyLys 386
1509 GTGCAACTGATGATCCGATATACGTTCAACCCGACAACTATTTTCG 1558
: : : : : : : : : : : : : : : : : : : : : : : : :
386 rLeuArgValLeuSerAspGlyLThrAlaArgGlyLThrIleLeuAsn.... 401
1559 GCTTTCGCGGCGACGCTTGGATTTAAACGGCATTCGCTTGTTCAC 1608
: : : : : : : : : : : : : : : : : : : : : : : : :
402 ....AsnSerGlyArgGlnAsnValSerAsnGlnGlyLysIleSerThrAsn 416
1609 CGTATTCAAATACCGATGAAAGGGCGATGATTCNCATATATGCGCAC 1658
417 AlaMetIleAsnThrGlnGlyLysGlnIleIleIyrSerAspGlyLys 433
1659 AACACATCCACCGCTTACCATTCAGCG.....AATGAAGTA 1696
: : : : : : : : : : : : : : : : : : : : : : : : :
433 aThrAlaAlaIleValAsnThrSerGlyPheGlnArgIleAsnSerGly 450
1697 TTACACACCG..... 1707
450 LThrAlaPheValGlnAsnSerValValIleThrArgThrValSerSer 466
1708 AGTGTAAGATATATCAATAGACTTAATTACAGC.....AAGAATATTC 1751
467 AlaAlaLysPhePheAspAlaGlnValIleIyrSerGlyLysGlnIle 483
1752 CTAC.....AACGTTGCTTGGCAGAAAGATAC..... 1782
: : : : : : : : : : : : : : : : : : : : : : : : :
483 lThrLeuThrArgGlyLThrIleIyrSerAsnPheLeuThrAlaValTrps 500
1783 .....ACCAAAACGACGCGCGG 1800
500 erMetPheProGlyLThrAlaSerGlyAlaAsnValAsnLeuSerGlyArg 516
1801 CTCAAC..... 1806
517 LeuAsnAlaPheAlaGlyLysAsnValValGlyLThrIleLeuAsnGlnIle 533
1807 .....CTGTATACAGCCCGCGCA..... 1827
533 yArgGlnIleIyrSerGlyLThrAlaThrSerThrValGlyLysAsn 550

```

```
1828 ..GAAGACGGCACCCNGCTGCTTTCCGGCGGAACAATTTAAACGGCAAC 1875
      |||
      |||
550 smgluylarglulrvalleuserglylylthr...Aspglythr 565
1876 ATGACGAAACAACGCAACTGTTTTCAGC.....GGCAGACCGAC 1919
      :
      :
566 ValLeuAsnSerIyglYleuGlnAlaValSerSerIyglYleuAlaSe 582
1920 ACCGCAAGCCTTACATCATTTAGGAAGCGGTGTCAAAAATGGAAGTA 1969
      :
      :
582 rAlatThrValIleAsnGluGlyAla..... 591
1970 TCCCAACAGAGAAATCGTGTGGACACAGCTGATCNACCGCAGCTTT 2019
      :
      :
592 .....GlnPheValTyraSpGlyGlyAlnValThrGlyThrAsn 604
2020 AAGCGGAAATTTCCATATTCAGGGCGGCGACGCGGTATTTCCCGCA 2069
      :
      :
605 IleLysAsnGlyGlyThrIleArgValAspSerGlyAlaSerAlaLeuAs 621
2070 TGTTCGCAAGTGGAGCGGATTGNCATTGAGCAAT..... 2106
      :
      :
621 nIleAlaLeuSerSerGlyAlaSnleuPheThrSerThrGlyAlaThrL 638
2107 .....CACGCCAAGCAGTTTGTGTGCGCACCGCATCAAGC 2145
      :
      :
638 euProGluLeuThrThrMetAlaAlaLeuSerValSerGlnAsnHisAla 654
2146 CATCAATCTGT..... 2157
655 SerAsnIleValleuGlnAsnGlyLeuLeuArgValThrSerOlygl 671
2158 .ACACGTTGGAAGTGCACNGTGTGACAATTTGTGCAANAANAACATTA 2206
      |||
      |||
671 yThrAlaThrAsp...ThrThrValAsnSerAlaGlyArgLeuArgIleA 687
2207 CCGACGATAAAGTGTGCTTCACTGACTAAGACAGACNTNAGCGGCANT 2256
      |||
      |||
687 spAspGlyGlyThrIleAsnGlyThrThrIleAsnAlaSpGlyIle 703
2257 GTNAGNCTNCCNATNACGNTNNTNNAANCNCNGCGGCTGNCNCCT 2306
      |||
      |||
704 Val.....AlaGlyThrAsnIleGlnAsnAspGlyAsnPh 715
2307 NAANGCAATCTTACTGCAATGCGCATACGTTATACATGACGACCA 2356
      :
      :
715 eIleLeuAsnLeuAlaGluAsnTyraSpPheGluThrGluLeuSerGlys 732
2357 AGGCC.....ACCAAAACGGCAAC 2376
732 ergIyValleuValLysAspAsnThrGlyIleMetThrTyraGlyThr 748
2377 CTYAGCCTC.....GTGGGCAATGCCCAAGCAACATTTAA 2411
      |||
      |||
749 LeuThrGlnAlaGlnGlyAlaAsnValLysAsnGlyGlyIleIlePheAs 765
2412 TCAGCCACATTAACGCAACNCATCGGNTCGGGCAATGCTTCATTTA 2461
      :
      :
765 pSerAlaValAlaAsnAlaSpMetAlaValAsnGlnAsnAlaTyrllea 782
2462 ATCTAAGCAACAAGCGGACAAACGGCAGCTGACGCTTCCGACAAAC 2511
      :
      :
782 snIleSerAspGlnAlaThrIleAsnGlySer.....ValAsnAsnAsn 796
2512 GCTAAGCAACGTAAGCATTCGCACTCAACGCAATGTCCTCCTAGC 2561
      :
      :
797 GlySerIleValIleAsnAsnSerIleIleAsnGlyAsnIle..... 810
2562 CGATAAGCAGATTCATTTTGAACAACGCGCTTTACCGGACAACTCA 2611
      :
      :
811 .....ThrAsnAspAlaA 815
2612 GCGGAGCAAGGANAACGATTAACACTTAAAGACAGCAAGTGAACGCTG 2661

      :
      :
815 spleuSerPheGlyThrAla.....LysLeu 823
2662 CCGTCAGCGACGAATTAGCAATTTAAACCTTGACACAGCCACCATTA 2711
      :
      :
824 LeuSerAlaThrValAlaSnGlySerLeu...ValAsnAsnLysAsnIleI 839
2712 ACTCAATCCGCCCTATCCGCACGATGCTGCGAGCGCCCAACCGGCAAG 2761
      :
      :
839 eLeuAsn..... 841
2762 TGTGACACAGCCCGCGCGCTTGCGGCGCTTCCTATTATCCGTTACA 2811
      :
      :
841 ..... 841
2812 CCGCCAACTTGGTAGAATCCGTTTCAACACGCTGACGTAACGGCA 2861
      :
      :
842 ...ProThrLys...GluSerAlaGlyAsnThrLeuThrValSer... 854
2862 ATTGAACGTCAGGAACATTCGCTTATGTCCGAACCTCTTCGGCTAC. 2910
      :
      :
855 ...AsnTyThrGlyThrProGlySerValIleSerLeuGlyGlyValL 870
2911 .....CGAAGCAACAATGAGCTGCGGCAAGTTCC 2943
870 euGluGlyAspAsnSerLeuThrAspArgLeuValValGlyAsnThr 886
2944 GAAGGNACTTACCTTGGCG...GTCAACAAATACGGCAACGAACCCGT 2990
      :
      :
887 SerGlyGlnSerAspIleValTyraValAlaSnGluAspGlySerGlygl 903
2991 AAGCCTGATCAATTAAGCGTAGTGAAGGAAGGAAGCAACCAACCGCTG 3040
      :
      :
903 nThrArgAspGlyIleAsnIleIleSerValGluGlyAsn..... 916
3041 CCGAAACCTTAATTTACCTTCGCAAAACGACAGCTGATGCGGCGCG 3090
      :
      :
917 ...SerAspAlaGluPheSerLeuLysAsn...ArgValAlaAlaGlyAla 931
3091 TGGCGTTACCACTCATCCGCAAAAGCGGAGTTCGCCCTGCATTAATCC 3140
      :
      :
932 TyraSpTyThrLeu...GlnLysGlyAsnGluSerGlyThrAspAsn.. 946
3141 GGTCAAGAACAAAGAGCTTCCGCAAACTCGGCAAGCAAGCAACCAAAA 3190
      :
      :
947 ...LysGlyTyThrLeuThrSerHisLeuProThrSerAspThrArgG 962
3191 AACAGCGCGGAAAAGACACAGCGGCAAGCCTTGACGCGCTGATTCGCGCC 3240
      :
      :
962 InTyraArgProGluAsnGlySerTyraIleThrAsnMetAlaLeuAla... 977
3241 GGGCGCGATGCGCGCGCAAAAGACAGAAAGCGTTGCCAAGCCGCGCGCN 3290
      :
      :
977 ..... 977
3291 GGCAGCGCGGGAATATGCGCATTTATGACGGGAGAGAGAAAAAAC 3340
      :
      :
978 .....AsnSerLeuPheLeuMetAspLeuAsnGluArgLysG 990
3341 GCGTCAGCGCGGATTAAGACAGCGCNTTGGCGAAACAGCGCAAGCGGAA 3390
      :
      :
990 InPheArgAlaMetSerAspAsn..... 997
3391 ACCGCGCGGNTTACCAACCGCTTCCCGCGCGCGCGCGCGCGCGGGA 3440
      :
      :
998 ThrGlnProGluSerAla..... 1003
3441 TTTCGCGCAACCGACGCCCAACCGCAACCTCAACCCCAACCGACAGCG 3490
      :
      :
1003 ..... 1003
3491 ACCTGATNACCGTTATGCAATAGCGGTTTGAGTAATTTTCCGCGCAG 3540
```

```

1003 ..... 1003
3541 CTCACAGGTTTGGCCGTACAGACGAATGACCGCGTGTTCGCGA 3590
1003 ..... 1003
3591 AGACGGCGCACGCGTGG.....ACAAGCAGATCCGCGNACA 3631
1004 .....SerValTrrMetLysIleThrGlyIleSerSerG 1016
3632 CCAAACTACTCCGTTGCGAAGATTTCGCGCTACCGCAACAAACGAC 3681
1016 Lys..... 1017
3682 CCGCGCAATGCGTATGAGAAAAACCTGGCGAGCGCGCGTGGCAT 3731
1018 .....LeuAsnAspGlyGln..... 1022
3732 CCGTGTTCGACAAACCGACGAAACACTTCGACGAGCGCATGCGCA 3781
1023 .....AsnLysThrThrAsnGlnPheIleAsnGlnLeuGly 1036
3782 ACTCGGACAGCGCTGCCCGCGCGCGTGTTCGCGCATGCGCATGCGC 3831
1036 LysPheLysPheHisAlaGlu.....GlnLeuGly 1047
3832 AGTTGACATCGGCGATCAGACGCGCGCGTGTTCGACGCGCGCATCT 3881
1048 AspPheThrLeuGlyIleMetLysGlyTyrAlaAsnAlaLysGlyLys 1064
3882 ATCAGAGCGCATCGAGGCAAAATCCGCGCGCGTGTTCATTCAGCGCA 3931
1064 rIleAsnTyrThrSerAsnLysAlaAlaArgAsnThrLeuAsp..... 1078
3932 TTCAGCAGATACCGCGCGTTCGCGCGATGCGCATGCAACCGTAC 3981
1079 .....GlyTyrSerValGlyValTyrGlyThrTrrPrrGlnAsnGly 1092
3982 ATCGGCGCAAGCGGTATTTCGTCGCAAAA.....GCGGATTACCGGTACGA 4028
1093 GlnAsnAlaThrGlyLeuPheAlaGlnTrrPrrMetGlnTyrAsnTrrPrr 1109
4029 AAACGTCAATATCGCCACCGCGTGTTCGTTCAACCGTACGCGN.... 4074
1109 eAsnAlaSerValLysGlyAspGlyLeuGlnGlnLysTyrAsnLeuA 1126
4074 ..... 4074
1126 snGlyLeuThrAlaSerAlaGlyGlyTyrAsnLeuAsnValHisThr 1142
4075 .....GCGGCAATTAAGGAGATTAATTCATTCGCAACCG..... 4107
1143 TrrPrrSerProGlnGlyIleThrGlyLysPrrPrrLeuGlnPrrHisIle 1159
4108 .....GCGCAACATNTTCATCAGCAGCTTATTTAGCTGTGCTATACCG 4153
1159 uGlnAlaValTrrMetGlyValTrrProAspPrrHis..... 1171
4154 ATGCCCGCTCGGCAAGTCCGACACGCGTCAATACCGCGTATTTGGT 4203
1172 .....GlnGluAspAsnGlyThrValVal 1179
4204 CAGGATTTCGCAAAACCGCG.....AGTGGCGA 4232
1180 GlnGlyAlaGlyLysAsnAsnIleGlnTrrLysAlaGlyTrrArgAlaIle 1196
4233 ATGGGCGCTAAACGCGCAATCAAGTTTCAGTGTGCTCCTTCACGCGTG 4282
1196 rTrrPrrValLysSer.....ThrLeuAspLysAspThrG 1208
4283 CCGCGCGCAAGGCGCACTGGAAGCGCA.....CAC 4317
1208 LysArgPrrPrrArgPrrTrrIleGlnAlaAsnTrrPrrIleHisAsnThrHis 1224

```

```

4318 AGCGCGGCATCAATTA 4335
::: |||::|::|::|
1225 GlupheGlyValLysMet 1230
seq_name: SwissProt_40:AMYH_YEAST
seq_documentation_block:
ID AMYH_YEAST STANDARD: PRT; 1367 AA.
AC P08640: P08068:
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucosylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
glucosylase) (1,4-alpha-D-glucan glucosylase).
GN STAI OR STA2 OR MAL5 OR YIR019C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrall B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE=87194600; PubMed=3106330;
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT STAI.",
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STA2 and STA genes
RT from Saccharomyces cerevisiae.",
RL FEBS Lett. 239:179-184(1988).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1- SIMILARITY: TO S.POMBE SPC215.13.
CC -1- SIMILARITY: SOME, TO S.POMBE SPC285.13C.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z38061; CAAB6176.1; -
DR EMBL: M16164; AAA35014.1; -
DR EMBL: M16165; AAA35015.1; -
DR EMBL: X13857; CAA32069.1; -
DR PIR: B26877; B26877.
DR PIR: B26877; B26877.
DR PIR: S48476; S48476.
DR SDD: S0001458; MDC1.
KW Hydrolyase; Glycosylase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21
FT CHAIN 22 1367
FT DOMAIN 210 1367
FT CARBOHYD 817 817
POTENTIAL.
GLUCOSYLASE S1/S2.
SER/THR-RICH.
N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1367 AA: 136110 MW: 91C00E2DBD61AA9D CRC64;

## alignment\_scores:

Quality: 206.00 Length: 754  
Ratio: 0.564 Gaps: 32  
Percent Similarity: 48.408 Percent Identity: 20.557

## alignment\_block:

US-09-303-518D-651 x AMYH\_YEAST ..

Align seg 1/1 to: AMYH\_YEAST from: 1 to: 1367

```
2134 CCGCATCAAGGCGCATGACATGCTGACGCTGCGACGCTGACGCTGAC 2163
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
268 ProHSHLSAspThrThrProCysThrLysLysThrThrThrSeryl 304
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2184 AATGTGTGCAANAANAACATTCGACGATTAAGTATGCTTCATTTGA 2233
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
304 sThrcysThrLysLysThrThrThrProValProThrProSerSert 321
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2234 CTAAGACAGACNTNAGCGGCGANTGAGNCCNCCNATNAGGNTNTNA 2283
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
321 hrThgluSerSerAlaProValProThrProSerSerThrThr 337
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2284 AANCTCNGGCGCNTGNCNCACTNANAGCAATCTTAGTCAATGGGGA 2333
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
337 rgluSerSer.....AlaProValThrS 346
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2334 TACACGTTATACAGTACAGCCACAGCCCA.....CCCAAAAGCGCA 2374
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
346 erSerThrThrgluSerSerAlaProValProThrProSerSert 362
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2375 ACCTAGCGCTGTGGGCAATGCCAGACACATTTAATCAAGCCACATTA 2424
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
363 ThrThrgluSerSerAlaPro..... 370
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2425 AACGGCAACNCATCGGNTTCGGCAATGCTTCATTTAATCAAGCAAA 2474
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
371 .....ValThrSerSerThrT 376
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2475 CGCGGCAACAAAGCGAGTGTGACGCTTCCGACACCGCTAAGCAAG 2524
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
376 hrgluSerSerAla.....ProValThrSerSerThrThr 388
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2525 TAAGCATTCGCGACTCAAGGCAATGCTCCCTAGCCGATAGGCACTA 2574
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
389 .....gluSerSerAlaProValProThrProSerSe 400
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2575 TTCCATTTTGAAGAGCGCGCTTACCGGACACTGACGCGCAGCAAGA 2624
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
400 rSerThrThrgluSerSerAlaProValThrSerSerThrThrgluS 417
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2625 NACGACATTACCTTAAAGACACAGCAATGAGCGCTGCCGTCAGCAGG 2674
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
417 erSerSer..... 419
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2675 AATTAGGCAATTTAAACCTTGACAGAGCGCAACATTACACTCAATTCG 2724
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
420 .....AlaProValThrSerSerThrThrgluSerSerAlaProVa 434
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2725 TATGCGCAGATGCTGACGCGCGCAAA...CCGCGAGNTGTGACAGAC 2771
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
434 lThrSerSerThrThrgluSerSerAlaProValThrSerSerThrT 451
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2772 GCCCGCGCGCGCTGCGCGCTTCCTAATTAATCCGTTACACGCCAACTT 2821
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
451 hrgluSerSerAlaProValPro..... 459
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2822 CGGTAGAAATCCGTTTCAACACGCTGACGTAACGGCAATTAAGACNGT 2871
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
460 .....ThrProSerSerThr 465
```

```
2872 CAAGAACATTCGCGCTTATGTGCGAAGCTTTCGCTACGGAAGCGACA 2921
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
465 rThrgluSerSerAlaProValThrSerSerThrThrgluSerSers 482
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2922 ATTTAGAGCTGGCGGAAGTTCGAGAGNACTTACCTTGCGGCGTCAACA 2971
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
482 er...AlaProValProThrProSerSerSerThrThrgluSerSer 497
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2972 ATACCGGCAAGAACCCGTAAGCCTCGATCAATTTGACGCTAGTGAAGG 3021
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
498 AlaProValThrSer..... 502
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3022 AAGACACAAACACCGCTGTCCGAAAACCTTAATTTACCTGCAAAACA 3071
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
503 SerThrThrgluSerSerAlaProValProThrProSerSert 519
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3072 ACACG...TCGATGCGCGCGCTGCGGTTACCACTCATCCGCAAGACG 3118
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
519 hrThrgluSerSerSerAlaProAlaProThrProSerSerThrThr 535
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3119 GCGAGTTCCGCTGCAATATCCGTCAAAGACAAAGACGCTTCCGACAA 3168
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
536 GluSerSerAlaProValThrSerSerThr.....Thrgl 548
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3169 CTGCGGCAAGGCAAGACGCAAAACAGCGGCAAAAGACAGCGCAAG 3218
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
548 uSerSerSerAlaProValProThrProSerSerThrThrThrgluSers 565
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3219 CATTGACGCGCTGATGTCGCGCGCGCGATGCGCGCGCAAAAGACAGAA 3268
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
565 erSerThr.....ProValThrSerSerThrThrThrgluSerSer 577
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3269 GCGTTGCCAAGCGCGCGCGCGCGAGGAGGGAATGTCGGCATTAAG 3318
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
578 SerAlaPro..... 580
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3319 CAGCGGAGGAGAGAGAAAAACGGGTGACGCGGATTAAGACAGCGCNTT 3368
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
581 .....ValProThrPro... 584
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3369 GCGGAAACAGCGGAGCGGAAACCGCGCGGNTTACACCGCTTCCGCC 3418
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
585 .....SerSerSerThrThrgluSerSerSerAlaProValProThrPro 599
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3419 .....GCGCGCGCGCGCGCGCGCGGATTTGCCGCA 3450
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
600 SerSerSerThrThrThrgluSerSerSerAlaProAlaPro..... 612
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3451 CCGCAGCGCCCAACCGCAACCTCAACCCCAACGCGCGGACCTGATNAG 3500
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
613 .....ThrProSerSerSerThrThrThrgluSers 622
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3501 CCGTTATGCGCAATAGCGGTTGAGTAATTTCCGCGCAGCTCAACAGCG 3550
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
622 erSerAlaPro.....ValThrSerSerThrThrThrgluSerSer 634
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3551 TTTTCCCGCTAGCAGAGCAATTTGACCGGCTGTTCCGAGACGCGCGC 3600
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
635 SerAlaProValProThr.....ProSerSerSerThr 645
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3601 AACGNGTTTGGACAGACNGCATCCGAGACACA...AACACTACCGGTC 3647
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
645 rThrThrgluSerSerSerAlaProAlaProThrProSerSerSerThrThr 662
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3648 GCAAGATTTCCGCGCTTACCGCGCAACAAACGAGCTGCGGCAAAATG 3697
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
662 luserSerSerAlaProVal.....ProThrProSerSerSerThr 675
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3698 TGCAGAAAAACCTGCGGACAGCGCGCGGTGCGATCTGTTTCGACAA 3747
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
676 ThrThrgluSerSerSerAlaProValThrSerSerThrThrThrgluSerSe 692
```





```

802 GACAAACA.....AACATTAATGCTGCTCAACGAGTTTTCACAAAC 845
    ||| ||| :||| :|||
303 AsphthrtlaalaalaglyLysleuIlelaasnly.....GlyAl 317
846 CGGCTACCTTATTCCGGCAGGAAACGGTTTCCAGCTGATCCGCAAG 895
    :||| :||| :|||
317 aalaasnalaValIleGlyThrAspAsnGlyAlagly.....ArgAlaA 332
896 ATGCTTACGATGACATTTTACAGAGCGGTACACATACCGTCTNTTT 945
    ||| :||| :|||
332 laGlyPheIleValSerValAspAsnGlyAlaIleThrIleSerGly 348
946 GAACCGCGCAGTACGACATTTTCTTACATCCAAACAACAGCT.. 993
    :||| :||| :|||
349 GlNValTyralaLysAspIleValIleGlnSerIleAsnIleGlyGly 365
994 .....ACGGTACGCTTACGAAACCAACG 1018
365 nValThrPheGlnHisLeuValaLysPValGlyLeuGlyGlyThrAsn 382
1019 AAAGGTTCATCCAAAGCTTAAGTACAGACAGTCCGACTTTTGAC 1068
    ||| :||| :|||
382 heLysThrAlaAspSerLysValIleIleThrGlu.....Asn 394
1069 GAATCTTTGAATGAACGTAT..... 1089
395 AlaserPheGlySerThrAspPheGlyAsnLeuAlaValGlnIleVala 411
1090 .....AAGAACCATTTACGGCGAGGGGGGTATATCAGTACCGTCCAA 1135
    :||| :||| :|||
411 lProAsnLysIleLeuThrGlyAsnPheIleGlyAspAlaLys.... 426
1136 GGTAAACAACAGGTGAACCTTCTTTATCGATTAC...GGCAGCGC 1182
    ||| :||| :|||
427 .....AsnAsnGlyAsnThrAlaGlyAlaIleThrPheAsnAlaAsnGly 441
1183 AAATCATCTTATCAAC.....AACATCAA 1208
    ||| :||| :|||
442 ThrLeuValSerGlyAsnThrAspProAsnIleValaValThrAsnIle 458
1209 C.....CAAGCGCGGCGGTGTGTTGTAAGGTATTTTA 1246
    :||| :||| :|||
458 sAlaIleGlyValGlnGlyAlaGlyIleVal..... 468
1247 CGGTCTCGCTGAACAAACGAAACGTGCGAGCGCGGCTCATATC 1296
    ||| :||| :|||
469 .....GlnLeuSerGlyIleHisGly 475
1297 AGTGA.....GACAG 1307
    :||| :||| :|||
476 lAlaGlnLeuAlaGlyLeuGlyAsnAlaGlySerIlePheLysLeuAlaAsp 492
1308 TACGTTACTTGAAGTAAACGGCGGTGCAACGACCGCTGCCAAA 1357
    :||| :||| :|||
492 yThrVal.....IleAsnGlyProValAsn..... 500
1358 TCGCAAAAGCAGCGCTGACGCTTACGCAAAAGGGAACCAAGGCTCG 1407
    ||| :||| :|||
501 .....GlnAsnProLeuValaLysAsnAsnAla 509
1408 ATCAGCGTGGC.....GACGATACGATCTTTGATCGATCAGA 1445
    :||| :||| :|||
510 LeuAlaIleGlySerIleGlnLeuValaLysPValIleIleThrGlyAs 526
1446 GGCAGAGCATAAAGCAAAACAGCTTATAGTGAATCGGCTGTC 1494
    :||| :||| :|||
526 PileGlyAsnGlyAlaValaLysAlaIleGlnAspIleThrLeuAla 543
1495 .....AGCGCAGGCGTACGCTCACTG 1518
543 snaAspAlaSerLysIleLeuThrLeuSerGlyAlaAsnIleIleGlyAla 559
1519 AATGCCGATAATCAAGTTCACCCGACAAACTATTTCCGCTTCCGG 1568

```

```

560 AsnAlaGlyGly.....AlaIleHisPheGlnAlaAsnGly 571
1569 CGGACTTTTGATTTA.....AACGGCATTCGCTTTGTTCC 1606
    ||| :||| :|||
571 yGlyThrIleGlnLeuThrSerThrGlnAsnAsnIleLeuValaAspPhe 588
1607 ACCGATTCAAATACGATGAAGGGCGATGATTCNCATCATTAATGCC 1656
588 sPheAspValThrThrAspGlnThrGlyValaLysAlaSerSerLeu 604
1657 ACAACAATCCACCGCTTACATTCAGGAAATGAATATTACACAAAC 1706
    ||| :||| :|||
605 ThrAsnAsnGlnThrLeuThrIleAsnGlySerIleGlyThrIleGlyAl 621
1707 GAGTGTAAATATCATATGACTTATATTCAGCAAA..... 1743
    :||| :||| :|||
621 aAsnThrLysThrLeuGlyArgPheAsnValaGlySerSerLysThrIle 638
1744 .....GAATTCGCTCAACGCTTGGTTGGCAGAAAGAT... 1779
638 euAsnAlaGlyAspValaIleAsnGlnLeuValaMetGlnAsnAspGly 654
1780 .....ACGACCAAAACG..... 1791
655 SerValHisLeuThrHisAsnThrTyrlleuIleThrLysThrIleAsnAl 671
1792 .....AACGGCGGCTCAACCTTGTATACAGCCCGCGCAGAAAGAC 1834
    :||| :||| :|||
671 aAlaAsnGlnGlyLysIleIleValaAlaAlaAspProIleAsnThrAsp 687
1835 GCACCCGCTGCTTCCGGCGCAACAATTA..... 1866
    :||| :||| :|||
688 .....ThrIleAlaAlaAspGlyThrAsnLeuGlySerAlaGlySerPro 702
1867 .....ACGGC..... 1872
703 LeuSerAsnIleHisPheAlaThrLysAlaAlaAsnGlyAspSerIle 719
1873 .....AACATCAGCAACAA 1888
719 uHisIleGlyLysGlyValaAsnLeuTyrlaAsnAsnIleThrThrAla 736
1889 AC.....GGCAACGTGTTTTCAGCGGAGACGACACCGCAGCGC 1929
    :||| :||| :|||
736 sPAlaAsnValaGlySerLeuHisPheArgSerGlyGlyThrSerIleVal 752
1930 TACATCATTTAGGAAGCGGTGCTCAAAATGGAAGTATCCACACAG 1979
    :||| :||| :|||
753 SerGlyThrValGlyGly.....GlnGlnGlyLeuLysLeuAs 765
1980 AGAAATCGTGTGGACACAGCAGTCAACCGGACGTTTAAAGCGGAA 2029
    :||| :||| :|||
765 naSnLeuIleLeuAspAsnGlyThrThrValaLysPheLeuGlyAspIle 782
2030 ATTTCATATTCAGGCGGAGCGGATTTCCGCAATGTGCCAA 2079
    :||| :||| :|||
782 hrPheAsn.....GlyGlyThrLysIleGlnGlyLysSerIleLeuGln 796
2080 GTGGAAGCGATTCATTTGACAAATCAGCCCAA..... 2115
797 lIleSerSerAsn...TyrlleThrAspHisIleGlnSerAlaAspAsnTh 812
2115 ..... 2115
812 rGlyThrLeuGlnPheValaSnThrAspProIleThrValThrLeuAsn 829
2116 .....CCAGTTTGTGTGTCACCGCATCAAAAGCATCAATCTGT 2157
    ||| :||| :|||
829 ySclnGlyAlaTyrlPheGlyValaLeuLysGlnValaMetValaSerGlyPro 845
2158 ACAGTTTCGAGCTGCACNGCTCTGACAAATGTGTGGAAMAANCAATTC 2207
    :||| :||| :|||

```

```

846 GlyAsnIleAlaPheAsnGluIleGlyAsnGlyValAlaHisAlaIleAl 862
2208 CGACGATAAAGTG..... 2220
862 aValAspSerIleSerPheGluAsnAlaSerIleuGlyAlaSerLeuPheL 879
2221 .....ATTGCTTCATTGACAGACAGACACNTN 2247
879 euleuSerGlyThrProLeuAspValIleuThrIleGlySerThrValGly 895
2248 AGCGGCAANTGANGCTNNCCNATNACGNTNNTTAAANCCTCNGCGCN 2297
896 AsnGlyThrValAlaAspAsnPheAsnAlaProIleLeuValIleSerGly 912
2298 TGCNNACTNANGGCAAT..... 2316
912 eAspSerMetIleAsnAsnGlyGlnValIleGlyAspGlnLysAsnIleI 929
2317 .....CTTAGTCAAAATGGGATACAGTTATACAGTCAGCCACAGCC 2361
929 lAlaLeuSerIleuGlySerAspAsnSerIleThrValAsnSerAsnThr 945
2362 .....ACCCAAACGGCAACCTTAGCT 2384
946 LeuThrAlaGlyIleArgThrThrIleAsnGlnGlyThrValThrIle 962
2385 CGTGGGC..... 2391
962 uSerGlyGlyIleProAsnAsnProGlyThrIleIleGlyLeuGlyLeuG 979
2392 ..AATGCCAAGCACATTAAATCAAGCCACATTAACGGCAACATCAG 2439
979 lAsnGlyAspProLysIleuLysGlnValThrPheThrAspTyrAsn 995
2440 GNTTGGGCAATGCTTCAATTATCTAAGCAACAGCCGCAACAAAGG 2489
996 AsnLeuGlySerIle.....IleAlaThrAsnValThrIleAsnAs 1009
2490 CAGTCTGAGCTT..... 2502
1009 pAspValThrLeuThrThrGlyGlyIleAlaGlyThrAspPheAspGlyL 1026
2503 .....TCCGACAACGCTAAGGCAACGTA..... 2526
1026 yAlleThrLeuGlySerIleAsnGlyAsnAlaAsnValLysPheValAsp 1042
2527 .....AGCCAT..... 2532
1043 ArgThrPheSerHisProThrSerMetIleValSerThrLysAlaAsnG 1059
2533 .....TCCGACTCAACGGCAATGTCCTCCTAG 2560
1059 nGlyThrValThrTyrLeuGlyAsnAlaLeuValGlyAsnIleGlySers 1076
2561 CGGATAGGCAATTCATTGTTGAAAACAGCCGCTTACGGCAACACTC 2610
1076 eAspIleProVal.....AlaSerValArgPheThrGlyAsnAsp 1089
2611 AGCGGCAAGCAAGANACGCA..... 2631
1090 SerGlyValGlyLeuGlnGlyAsnIleHisSerGlnAsnIleAspPheG 1106
2632 .....TTACACTTAAAGACAGCAATGACGCTGCGCTCAGCA 2671
1106 yThrTyrAsnLeuThrIleLeuAsnSerAspValIleLeuGlyGlyGlyT 1123
2672 CGGAATTAGCAATTTAAACCTTGAC..... 2697
1123 hThrAlaIleAsnGlyGluIleAspLeuThrAsnAsnLeuIlePhe 1139
2698 ...AAGCGACCATTAACACTC...AATCCGCTATCGCAGCAGAGCGC 2741
1140 AlaAsnGlyThrSerThrPglYAsnAsnThrSerLeuSerThrThrLe 1156

```

```

2742 AGCGCGCAAAACGGCAGANGTGCAGACAGCCGCGCGCTTCGGCC 2791
1156 uAsnValSerAsnGlyAsnValGlyGlnIleValIleAlaGlnGlyAlaG 1173
2792 GTTCCCTAATATCCGTTACACCCGCAACTTCGGTAGAATCCGTTTCAAC 2841
1173 lValAsnAlaThrThrThrGlyThrThrIleLysIleGlnAspAsn 1189
2842 .....ACGCTGACGGTAAACGGCAATT 2864
1190 AlaAsnAlaAsnPheSerGlyThrGlnThrThrLeu..... 1202
2865 GAACGTCAGAGCAATTCGCTTATGCGAATTCCTGCTACCGAA 2914
1203 ...IleGlnGlyValAlaArgPheAsnGlyThrLeuGlyAlaProAsn 1218
2915 GCGAC.....AATTTGAAGCTGGCGGAAGT 2940
1218 hAspValThrGlyAsnAsnIlePheValLysTyrGlnLeuIleArgAsp 1234
2941 TCCGAAAGNACTTACACTTCGGCGGTCAACATACCGCAGACACCCCT 2990
1235 AlaAsnGlnAspTyrValLeuThrArgThrAsnAsp..... 1246
2991 AAGCCTGATCAATGACGGTAGTGAAGGAAAGAC.....ACA 3031
1247 .ValLeuAsnValValThrThrAlaValAlaGlyAsnSerAlaIleAlaAsn 1263
3032 AACCG...CTGCCGAATAACCTTAATTCACCTCGCAAAACGACAGTC 3078
1263 lApGlyValAlaHisGlnAsnIleAlaIleCysLeu.....GluSerThr 1277
3079 GATCGCGCGCGTGGCGTTCACCACTCATCCGCAAGAGCGGAG..... 3123
1278 AspThrAlaIleThrAsnAsnMetLeuAlaLysAspSerSerAspVa 1294
3123 ..... 3123
1294 lAlaThrPheIleGlyAlaIleAlaThrAspThrGlyAlaAlaValAla 1311
3124 ..TTCCGCTGCGATATCCGCTCAAGACAAAG..CTTTCGCAAAA 3168
1311 hValAlaLeuAsnAspThrGlnLysThrGlnAspLeuGlyAsnArg 1327
3169 CTGCGCAAGGAGAGCAAAACAGCGGAA...AAGACAAACGGCGCA 3215
1328 LeuGlyAlaLeuArgTyrLeuSerAsnSerGlnThrAlaAspValGly 1344
3216 AAGCCTTGACGCGCTGATTGCGCGCGCGCATGCCGCGGAAAGACAG 3265
1344 ySerGlnThrGlyAlaValSerSerGlyAspGluAlaIleAspGlnVal 1361
3266 AA.....AGCGTTCGCAACCGCGCGG 3288
1361 eTyrGlyValAlaThrAlaLysProPheTyrAsnIleAlaGlnAspLys 1377
3289 CNGCGCAGGC 3297
1378 LysGlyGly 1380

```

seq\_name: SwissProt\_40:OMPb\_RICCN

```

seq_documentation_block:
ID OMPb_RICCN STRANARD: PRT: 1655 AA.
AC Q9KRA3; Q9KK98; Q9XC45;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Seas) (TompB)
DE (Tomp B) (Contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).

```

GN OMPB OR RCI085.  
 OS Rickettsia conorii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 ON NCBI\_TaxId=781;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Mallish 7;  
 RA MEDLINE=21442074; PubMed=11557893;  
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,  
 RA Samsen D., Roux V., Cossart P., Weissensbach J., Claverie J.-M.,  
 RA Raoult D.;  
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";  
 RL Science 293:2093-2098(2001).  
 RN [2]  
 RP SEQUENCE OF 33-1649 FROM N.A.  
 RC STRAIN-Indian tick typhus, and Mallish 7;  
 RX MEDLINE=20393643; PubMed=10939649;  
 RX Roux V., Raoult D.;  
 RT "Phylogenetic analysis of members of the genus Rickettsia using the  
 RT gene coding the outer-membrane protein ompB (ompB)".  
 RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).  
 RN [3]  
 RP SEQUENCE OF 353-1655 FROM N.A.  
 RC STRAIN-Mallish 7;  
 RA Stenos J., Walker D.;  
 RT "The rickettsial outer membrane protein A and B genes of Rickettsia  
 RT australis, the most divergent rickettsia of the spotted fever group.";  
 RL Submitted (MAY-1999) to the EMBL/Genbank/DDAJ databases.  
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL FACTOR WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY  
 CC SIMILARITY).  
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR  
 CC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
 CC LAYER WITH HEXAGONAL SYMMETRY (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AE008659; AAL03623.1; -  
 DR EMBL: AF123721; AAF34124.1; -  
 DR EMBL: AF123726; AAF34129.1; -  
 DR EMBL: AF149110; AAD39533.1; -  
 DR InterPro: IPR003858; romPA\_ompB.  
 DR Pfam: PF02708; romPA\_ompB.1  
 KW Antigen; S-layer; Cell wall; Complete proteome.  
 FT CHAIN 1 1334 120 KDA SURFACE-EXPOSED PROTEIN.  
 FT CHAIN 1 1335 1655 32 KDA BETA PEPTIDE.  
 FT VARIANT 61 61 P -> A (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANT 75 75 G -> S (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANT 78 78 K -> N (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANT 251 251 V -> A (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANT 413 413 N -> D (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANT 959 959 I -> V (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANT 988 988 A -> T (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANT 1139 1139 R -> L (IN STRAIN INDIAN TICK TYPHUS).  
 FT CONFLICT 353 354 KD -> GH (IN REF. 3).  
 FT CONFLICT 776 776 F -> S (IN REF. 3).  
 FT CONFLICT 1159 1159 E -> D (IN REF. 3).  
 FT CONFLICT 1177 1177 G -> S (IN REF. 3).  
 FT CONFLICT 1192 1192 H -> R (IN REF. 3).  
 FT CONFLICT 1492 1492 H -> R (IN REF. 3).  
 SQ SEQUENCE 1655 AA; 168342 MW; E49E19377D5FCB37 CRC64;

alignment\_scores:

Quality: 203.00 Length: 1500  
 Ratio: 0.288 Gaps: 79  
 Percent Similarity: 46.933 Percent Identity: 19.400  
 alignment\_block:  
 US-09-303-518D-651 x OMPB\_RICCN ..  
 Align seg 1/1 to: OMPB\_RICCN from: 1 to: 1655

```

178 AATAAGGCAAGTTGCTCCGCGGGAAGATATTGAGTTTACACAA 227
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
339 ASPASngLYLys.....ValAtrrIleaspgLYGlnValTlyAlaly 353
228 AAAAGGGGAGTTGTCGCCAATCAATGACAAAAGCCGATGATTGAT 277
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
353 s.....AspmetValIleGlnSerIleasnAlaValGlyGlnValAsnp 368
278 TTTCT.....GTGTGTCGCGTACGCGCGTGGCGGCGATTGGTGGC 318
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
368 hearghisIleValaspaIglyThrAspGlyThrThrAlabheysthr 384
319 GATCATATATTGTGAGCGTGACATACGCGCGGTATACACAGGTGA 368
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
385 AlaAlerIysValAlaIleThrGlnAsnSerAsnpheGlyThrThrs 401
369 TTTTGTGCGGAAGGAAGNAATCCGATGACGCGTTTCTTACCAA 418
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
401 phtegly.....AsnleuAlaIleGlnI 409
419 TGTGTAAGAAAT.....AATTATAGCTTGCAAT 450
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
409 IeIleValProAsnthrmetThrIleuasngIYasnphethrIlyspala 425
451 TCACACCTTACAGCGC.....GATTATCATATGCG 482
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
426 SerAsnpGlyAsnthrAlaIleThrIleAspAlaAsn...GI 441
483 GCGTTGCGTAATTTGTCAGATGACAGACCTGCGAAGAGAGAGT 532
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
441 yThrIleuAlaSerIleAspAlaAsn...ValAlaValThrAsna 457
533 ACATGAGGGGGAATACCTATTCGAT.....AAAGAA 564
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
457 snIleThrAlleGlnAlaSerGlyAlaIleValIleGlnIleusergly 473
565 AATATCCCGAGCGTCCGATCGGCTGAGACACACACTATGGCGTTA 614
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
474 ThrIleAlaIleGlnIleuAsnthrIleuAsnAlaIleGlySerValPheIysIle 490
615 TGATGATGACAAA.....CAGCGGATTTATCTACTCGGCGCATGT 658
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
490 uAlaAspGlyThrIleValIleAsnGlyIysValasngIthrAla...L 505
659 TAATTGGCGCAATACACATATGCAAGGTTGGGGAATATATGCGTANT 708
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
505 euValIleGlyIleAlaIleuAlaIleGlyThrIleThrIleuaspgIysAla 521
709 AGTTTAGGCGGATGTCGCCATGCAACGACTATGGCCCTATGCCAT 758
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
522 ThrIleThrGlyAspIleIyasnAla..... 530
759 TGCAGGTGCGGAGGAGACAGCGGTTCCGCAATGTTATTATGACAAA 808
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
531 .GlyGlyAlaIleAlaIleGlnGlyIleThrIleu...AlaAsnAspAlat 546
809 CAACATATATATGCTGCTCAAGGAGTTTACAAACGCGCTACCCCTAT 858
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
546 hrIysThrIleuThrIleu..... 551
859 TCCGGAGGGAACGAGTTCCAGCTGATACGCAAGATGTTGTTACGA 908
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
552 .....GlyGlyAlaAsnIleIleGlyAlaAsn.GlyGlyThrI 564

```

909 TGACCTTTACAGG.....GGCATACACATACCGTCNTTTTGAC 945  
564 LeasnhcgnlaaasnlglyglthrIleUvLeuthSerthrGln 580  
950 CGCGCGTACGAGCACTTTTCTTTCATCCAAACACGATCGAGT 999  
581 AsnIleValasp..PheaspLeuAlaIleAlaThrspGln..ThrAlly 596  
1000 ACCGTA.....ACAGAACCAACGAAAGTNTCCATCCAAAGCT 1040  
597 ValValaspAlaSerSerLeuthrAsnAlaGlnThrLeuThrIleasnG 613  
1041 TAAAGTACAGACAGCTCCGACGTGTGGACAACTTTGATGAAGAACTGATA 1090  
613 YlYsIIleGlyThrVal..... 618  
1091 AAGAACCGATTACGGCGAGGGGATTAAATCAGTACCGTCAAGTTA 1140  
619 .....GlyAlaAsnAsnIleThrLeuGlyInphe 628  
1141 AACACGCGTGAACCTTCTTTATCGATTCAGGCAAC.....GG 1181  
629 AsnIIleGlySerSerIleThrValLeuSeraspGlyaspValAlaIleas 645  
1182 CAACCTCATCTTATCAAAACATCAACCAAGCGCGCGGCGGTGTGATT 1231  
645 ngIuIeuValIIleGlyAsnAsn.....GlyAlaValGlnp 657  
1232 TT.....GAA 1230  
657 hehAlaIleasnThrIleThrIleThrAllyThrThrAsnAlaIleGlyIn 673  
1237 GGTGATTTTACGGTCTCGCT..GAAACAACGAACGCTGGCAAGGCG 1283  
674 GlyIleIleIlePheAsnProValValAlaAsnAsnThrThrLeuAlaTh 690  
1284 GGGCGTTCATACGAGAACAGACAGTACCGCTTACTGGAAAGTAAC... 1329  
690 rGlyThrAsnLeuGlySerAlaThrAsnProLeuAlaGluIleAsnPhg 707  
1330 .....GGCGTGGCAAC..GACCGCGTCTCCAAATCGCAAAAGC 1368  
707 LysrIleGlyAlaIleAsnValaspThrValLeuAsnValGlyIleGly 723  
1369 .....ACGCTGACCGTTCAAGCCAAAGGAAACCAACGCTC 1406  
724 ValAsnLeuIleAlaThrAsnIleThrThrAspAlaAsnValGlySe 740  
1407 G.....ATCAGCGTGGGAGCGATCACTGTTTGGATCGACGAG 1450  
740 rPheIlePheAsnAlaGlyIleThrAsnIleValSerIleThrValIleG 757  
1451 ACGATTAAGGCAAAACAAAGCCTTACTAGAAATCGGCTGTCACGCG 1500  
757 LysInGInGlyAsnIle.....PheAsnThrValAlaIleAsnspasnGly 771  
1501 AGGGGTACGGTCACTGAATGCCATTAATCAGTTCAACCCGACAAACT 1556  
772 Thr...ThrValIlePheLeuGlyAsnAlaIleThrPheAsnGlyAsnThr 787  
1551 CATTTTCGGCTTTCGGCGGCGACGTTTGGATTAAACGGG..CATTCGC 1597  
787 rIleAlaAla.....AsnSerThrLeuGlnIleIleGlyIleAsnIleThr 802  
1598 TTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGCGCATGTGTCNAT 1644  
802 IaspPhe.....ValAlaSerAlaIleaspGlyThrGlyIleValGluPhe 816  
1648 CATAATGCCAACAACATCCACCGTTACCTTACAGGAGGAATAT 1697  
817 ValAsn.....ThrIleProIleThrValThrIleAsn..... 827  
1698 TACACAACGAGTGTAGAAATATCAATATGACTTAATATACAGCAAGAAA 1747

828	.....	:	:::	:	:	839
1748	TTGCCTACAAACGGTTGGTTGGCGAGAAAGATACGACCAAAACGACGG	LYSGINALALAPROVALASNALALEUYSGINT				1797
839	lthr.....	:	:	:	Valsergly	843
1798	CGGCTACACCTGTTTACCAGCCCGCGCAGAAAGCCGACCCNCGT					1847
844	ProGlyAsnValValleLeu.....	:	:	:	Il	852
1848	TTCCGGCGAGACAAATTTAAAGCGACATCAGCAACA.....					1887
852	eglyAsnAlaglyAsnTtyHISgLYAlaValthrSphrThrlealpheg	:	:	:		869
1888	.....	:	:	:		1926
869	lunserserleuglyAlaValalPheleuProArglylePro...	:	:	:		884
1927	GCTTACAAATCATTTAGAAC...	:	:	:		1947
885	...PheAsnspAlaglyAsnArgileProleuthrIleYsSerThVa	:	:	:		900
1948	..GGTGTGTCAAAATGTGAAGT.....	:	:	:		1968
900	lGlyAsnLysThrAlaThGlyPheAspValProSerValleValleug	:	:	:		917
1969	.....	:	:	:		2001
917	lyAlaspservalleValaspglyGlnvalIleGlyAspLinsnAsn	:	:	:		933
2002	TGATCNACCGCAGCTTTAAACGGCAAT.....	:	:	:	TT	2033
934	lleValglyLeuLylLeuLylserSpsAspIleleValasAlaTh	:	:	:		950
2034	CCATATTCAGGGGGCAGCGGTGATTCGCCGATGTGCCAAAGTG	:	:	:		2083
950	rThrleuTYAlaglyIleGlyThrIleAsnSnsnngIlyThrValT	:	:	:		967
2084	AAGCGATGNCATTTGAGCAATCAGCCCAAGAGTTTT.....	:	:	:		2124
967	hrleuserglyIleProAsnThrProGlyThrValtyGlyLeuGly	:	:	:		983
2125	..GGTGTGCGACCGCATCAAGCCATCAATCTGTACACGTGGAGCTG	:	:	:		2171
984	ThrglyIleGlyAlaSerLysPheLysGlnValThrPheThrThAspTy	:	:	:		1000
2172	GAGCGGTGACAAATGTGTGCAANMAANCAATTCAGCAGTAAAGCA	:	:	:		2223
1000	rasnasnleuglyAsn.....	:	:	:	IleI	1007
2222	TTGCTTCATGACTAAGACAGCANTNAGCGGAGTGTNAGCTNNCAAT	:	:	:		2271
1007	lealThrAsnAlaThrIleAsnAspglyValThrValThrThrGlyGly	:	:	:		1023
2272	NACGNTNNNTTMAANCTCMKNGGCGTGCNNNACATNMAAGCAATCTAG	:	:	:		2322
1024	lleAlaglyIleGlyPheAspglyLysIleThrleu...GlySerValas	:	:	:		1039
2322	TGCAATGGCGATACAGCTTATCA.....	:	:	:	GTCAGCCCAACGCGCA	2362
1039	nglyAsnGlyAsnValAlarPheValaspglylleuSerHisSerThr	:	:	:		1055
2363	CCCAAAAGCGCAACTTATCCTCGTCGAGCAATGCCCAAGCAACATTAAT	:	:	:		2412
1056	.....	:	:	:		1066
2413	CAAGCCACATTAAAGCGCAACNCATCGNTTGGCGCAATGCTTCATTTAA	:	:	:		2462
1065	.....	:	:	:		1076
2463	TTTAAGCAACACGCCGCGCAAAAGCGCAGTGTGACGCTTTCGCAACG	:	:	:		2512

```
1077 .....GlyAsnIleGlyAspSerAspThrP 1085
2513 CTAAGCAACGTAAGCCATTCGCACTCAACGCAATGTCTCCAGCC 2562
      |||:|||||
1085 roValAlaSerValArgPheThrGlySerAspGlyValaGlyLeu... 1100
2563 GATAAGCAGTATTCATTGAAAAACAGCCGCTTACCGCAACCTCAG 2612
      |||:|||||
1101 .....GlnGlyAsnIleTyrSerGlnValI 1109
2613 CGGCACACGAGANACGACTTACCTATAAAGACGCAATGACGCTGC 2662
      |||:|||||
1109 eAspPheGlyThrTyrAsnLeuGlyIleSerAsnSerValIleLeuG 1126
2663 CGTCAGCAGCAGGAATTA.....GGCAATTTAACCTTACACACGCCAC 2706
      |||:|||||
1126 LyGlyGlyThrThrAlaIleAsnGlyLysIleAsnLeuArgThrSnrhr 1142
2707 ATTACACTCAATTCGCTATGCGACGATGCTGACGGCGCAACCGG 2756
      |||:|||||
1143 LeuThrPhe.....AlaSerGlyThrSerThrTr 1152
2757 CAGNGTGTGACAGACGCGCGCGCGCTTCGCCGCTTCCCTATATCCG 2806
1152 pGlyAsnAsn..... 1155
2807 TTACACCGCCCACTTCGTGAGATCCGCTTCAACACCGTCAGGTA... 2853
      |||:|||||
1156 .....ThrSerIleGlyThr.....ThrLeuThrLeuAla 1165
2854 AACGGCAATTTGACNGTCAGGAACATTCGCTTATGTCGGAACCTT 2903
      |||:|||||
1166 AsnGlyAsnIle.....GlyAsnIleValIleLeu... 1175
2904 CGGCTACCGAAGCAGCAATTTGACCTGCGGAAGTCCGCAAGNACTT 2953
1176 .....GlnGlyAlaGlnValAsnAlaThrThrThrGlyThrT 1188
2954 ACACCTTGGCGGTCAACATACCGCAACGAAACCGCTGATCATCA 3003
      |||:|||||
1188 hrThrIleLysValGlnAspAsnAlaAsnAlaAsnPheSerGlyThrGln 1204
3004 ...TTGACGGTACTGGAAGG.....AAAGCAACAACCGCTG...TC 3041
      |||:|||||
1205 ThrTyrThrLeuIleGlnGlyAlaArgPheAsnGlyThrLeuGlyG 1221
3042 CGAAACCTTATTTACCCCTGCAAAAGACAGACGTGATCCGCGCGGT 3091
      |||:|||||
1221 yProAsnPheValIleThrGlySerAsnArgPheValAsnTyrGlyLeuI 1238
3092 GGCGT.....TACCAACTCATCCGCAAAAGACGGGAGTTTC 3126
      |||:|||||
1238 leArgAlaAlaAsnGlnAsnPyrValIleThrArgThrAsn..... 1251
3127 CGCGTCGATTAATCCGTCAAAGACAGAGCTTCCGCAAACTCGGCAA 3176
      |||:|||||
1252 AsnAlaGlnAsnValIleThrAsnAspIleAlaAsnSerSerPheGly 1268
3177 GGCAGAACCCCAAAAAACAGCGGCAAAAGACAGCGCAAGCCGTGACG 3226
      |||:|||||
1268 yAlaProGlyVal.....GlyGlnAsnValIleThrT 1278
3227 CGCTGATTTGGCGCGCGCGATGCGCGCAAAAGACAGAAAGGTTGCC 3276
      |||:|||||
1278 hrPheValAsnAlaThrAsnThrAlaIleTyrAsnAsnLeuLeuAla 1294
3277 GAACGCGCGCGCGCGAGCGGGGGAAT...GTCCGATTAATGACAGC 3323
      |||:|||||
1295 Lys.....AsnSerAlaAsnSerAlaAsnPheValGlyAlaIleValTh 1309
3324 GGAGAGAGAGAAAAACGGGTGACGGCGATTAAGACAGCGCNTTTGGCGA 3373
      |||:|||||
1309 rAspThrSerIleAlaIleThrAsnAlaGlnLeuAspValAlaLysAspI 1326
3374 AACAGCGGAGCGGAAAAACCGCGGNTACCAACCGCTTCCCGCGGCC 3423
      |||:|||||
1326 leGlnAlaGlnLeuGlyAsnArg..... 1333
3424 CGCNCGCCCGCGGATTTGCCCAACCGCAGCGCCCAACGCACTCA 3473
      |||:|||||
1334 LeuGlyAlaLeuArgTyrLeuGlyThrProGluThrAlaGlnMetAlaG 1350
3474 ACCCAACCGCAGCGCGACCTGATNAGCCGTTATGCCAATAGCGTTTGA 3523
      |||:|||||
1350 yPro..... 1351
3524 GTGAATTTCCGCGACGCTCAACAGCGTTTCCGCTACAGAGCAATG 3573
      |||:|||||
1352 ..GlnAlaGlyAlaIleProAlaIleValAlaAlaGlyAspGlnAlaVal 1367
3574 GACCGCGTGTTCGCGAAGACCGCGCAACGNGTTTGACAAAGCNGCAT 3623
      |||:|||||
1368 AspAsnVal.....AlaTyrGlyIleThrAlaLysProPh 1379
3624 CCGGNACACCAACACTAC...CGTTGCAAGATTTCCGCGCTACCGCC 3670
      |||:|||||
1379 eTyrThrAspAlaHisGlnSerLysGlyGlyLeuAlaGlyTyrLysA 1396
3671 AACAAACCGACCTCGCCAAATCGTATGCAGAAAAACCTCGCAGCGGG 3720
      |||:|||||
1396 lAlysThrThrGlyValIleGlyLeu..... 1405
3721 CGCGTCGCAATCCTGTTTTCGACAAACCGGACGAAACANTTCGACGA 3770
1406 .....As 1406
3771 CGGATCGGCACTCGCGACGCGCTTGCCACGGCGCGTTTGGGCAAT 3820
      |||:|||||
1406 pThrLeuAlaAsnAspAsnLeuMetIleGlyAlaAlaIle..... 1419
3821 ACGCATCGGAGGTTCGACATCGGCATTCAGCAACGGCGCGGTTTACG 3870
      |||:|||||
1420 ..GlyIleThrLysThrAspIle.....LysHisGlnAspTyrLys 1432
3871 AGCGGCANTCTTCAGACGCGCATCGAGGCAAAATCGCGCGCGGTGCT 3920
      |||:|||||
1433 LysGlyAspLysThrAspValAsnGlyPheSerPheSerLeu..... 1446
3921 GCATTACGGCATTCAGCAGCA.....TACCGCGCGGTTTCG 3958
      |||:|||||
1447 ...TyrGlyAlaGlnGlnLeuValLysAsnPhePheAlaGlnGlySerA 1462
3959 GCGGATTCGCGCATCGACCGTACATCGGC...GCAACGCGCTATTTCGTC 4005
      |||:|||||
1462 lAllePheSerLeuAsnGlnValLysAsnLysSerGlnArgTyrPhePhe 1478
4006 .....CAAAAACGCGATTACCGCTACGAGAAA 4031
      |||:|||||
1479 AspAlaAsnGlyAsnMetSerLysGlnIleAlaAlaGlyHisTyrAsps 1495
4032 CGTCAATATGCCACACCCCGCGTTCGTTCGTTCAACCGTACCGGCGGCA 4081
      |||:|||||
1495 nMetThr.....PheGlyLysAsnL 1502
4082 TTAAGCGAGATTTATTCATTAACCGCGCAACACATNTTCATCAGNCCT 4131
      |||:|||||
1502 euthrValGlyTyrAspTyrAsnAlaMetGlnGlyAlaLeuValThrPro 1518
4132 TATTTNAGCTGTCTATACCATGCGCGCTTCGGGCAAGTCGGAACAG 4181
      |||:|||||
1519 MetAlaGlyLeuSerTyrLeuLysSerSerAspGlnAsnTyrLysLunh 1535
4182 CGTCAATACCGCGMTATTTGGCTCAGGATTTGGCAAA...ACCCGAGTG 4228
      |||:|||||
1535 rGlyThrThrValAlaAsnLysGlnValAsnSerLysPheSerAspArgT 1552
```



```
220 euProAspSerGlyIleSerValAlaLeuGluAspAsnThr..... 233
1321 AAGTAAACGGGGTGGCAACGACCGCTGTCCAAAATCGGCAAGGAC 1370
233 ..... 233
1371 GCTGCACGTTCAAGCCAAAGGGAACCAAGGCTCGATCAGCTGGCG 1420
234 .....ValGInGlyIleValIlysAsnGInGlySerIleLysAlaGlyG 248
1421 ACGGTACAGTCAATTTGGATCAGCAGGACGATAAGGCAAAACAA 1470
248 IuIleThrLeu.....SerAlaIysGlyArgGlyGln 258
1471 GCCTTAGTGAATGCGCTGTGTCAGCGC..... 1500
259 AlaLeuAspSerLeuValMetAsnAsnGlyValLeuGluAlaThrLysVa 275
1501 .....AGGGTACGTCACACTGAATCCGATTAATCAGTTCAACC 1540
275 IserAsnLysAsnGlyLysValValLeuSerAlaAspAsn..... 288
1541 CCGAACAACCTCAATTTGCGCTTTCGCGGCGACGTTTGATTTAAACGG 1590
289 .....ValGluLeuAsnAsn 293
1591 CATTCGCTTTGCTTCACCGCTATTCAAAATCGATGAAGGGGCGATGAT 1640
294 GluSer.....AsnIleLysGlyIuI 301
1641 TGCNATCATATATGCCACACACATCCACGTTACCATTCACGAGGATG 1690
301 eValThrPheGlyAlaAspValThrSerAsnLysGluLeuLysAspAsnI 318
1691 AAGATATCAACACCGAGTGTGAATATCAAT.....AGACTTAAT 1734
318 IeLysIleThrSerLysThrLysSerValIThrSerProLysIleAsn 334
1735 TACAGC...AAGAATTTGCTTACACAGGTTGTTGGCGAGAAGAATAC 1781
335 PheThrGlyLysSerValAsnIleAsnGlyAsnPheGlyArgGluAspSe 351
1782 GACC.....AAACGACGCGCGGCTCAAC 1807
351 tThrThrHisTyrLysAspGluPheLysLysLeuAsnThrGluValAsnI 368
1808 TT.....GTTTACACGCGCGCGGACGACGACGCG 1836
368 IeAspValProAspAsnGluAsnIleArgIleAlaAspIleGluAspAsn 384
1837 ACCCGCTGCTTCCGGCGGAACA..... 1860
385 ThrGlyThrGlyThrThrGlyThrGlyThrSerSerPheIleGlnThrG 401
1861 .....AATTTAAACGCA 1873
401 yAlaLeuSerSerLeuLeuAlaAsnAsnGlyLysValAsnLeuGly 418
1874 AC..... 1875
418 snAsnValAsnIleSerGlyArgIleHisIleAspSerPheArgGlySer 434
1876 .....ATCAGCAACAAAC..... 1890
435 AspSerLeuLeuLysLeuThrAsnLysGlyHisIleAspIleAsnAla 451
1891 .....GGCAACGTTTTC..... 1905
451 aaSPiIleHisSerLysGlyArgLeuPhePheIleThrSerLeuGlnAsnG 468
1906 .....AGCGGACAGACGACGACGACGCGCTTCAATCATTTA 1941
468 IuGluAspPheLysSerAsnIleThrIleThrAspSerLysIleAsnLeu 484
1942 GGAACCGC..... 1950
485 GlyAsnGlyAlaMetGlyLeuGlyArgSerValAspGluLysAspTyrAs 501
1951 .....TGTCAAAATGGAAGT.....ATCC 1972
501 pasnaTgTTPGInLysThrGluGlySerGlnArgLysLysPheAspVal 518
1973 CACAGAGAAATCGTGGGACAAAC..... 1998
518 ysmetSerAsnValGluPheAsnGlnValAspAspValIleLeuAlaGly 534
1999 .....GACTGATCCMACCGACGTTTAAACGCA 2027
535 GlyPheGluLysValAsnLeuAspLysIleValAlaThrGlyGlnThr.. 550
2028 AATTTCCATATTCAGCGCGGCGGCGGTCGATTTCCGCATGTTGCCA 2077
551 .AsnPheTyrIleAspGly.....ValSerArgAsnGlyArgL 564
2078 AGTGGAA...GGCGATTGNCATTTGAGCAATCACGCCCAACA..... 2118
564 ySTyrgIuTyrGlyValLeuAspLeuAspLysArgThrGlnLeuSerGlu 580
2119 .....GTTTTGTTGTCGACCGCA 2138
581 LeuAsnGlnGlyArgArgArgTyrGlyTyrTyrAspLeuGluLeuAs 597
2139 TCAAGCATACATCTGTATACAGTTCGGACCTGACGTCGACAAAT 2188
597 pMetAsnArgAlaTyrLeuTyrArgPheAspLeuPheAlaThrLysAsnT 614
2189 GTGTCAANAANAACATTTACCGACGATAAGTG..... 2220
614 hrcLysArgSerThrIleLysAspThrGluIleAsnIleSerAsnSeran 630
2221 .....ATTGCTTCATGACTAAGAC 2240
631 IleAsnLeuLysAsnGlyPheValHisLeuLeuAlaGluLysIleLys 647
2241 NGACNTNAGC.....GGCANTGTNA 2260
647 uaSPaAsnSerLysIleAspIleThrPheAspLysAspAsnSerGlnAsPT 664
2261 GNCNTNCCNATNACGNTNNTTNAANCTCNCGG...CNTCCNMCACNTN 2307
664 hrcLeuAlaGlnThrAsnArgLeuGlyMetAsnGlyLysValSerMetIle 680
2308 AANGCAATCTTAGTCAATGAGCGATACAGCTTATACGTCAGCCACA 2357
681 AsnSerHisIleLysIleValAlaGlyAspGluLysGluGlyTleSerProh 697
2358 CGCCACCCAAACGGCACCTT.....AGCTCGTGGGCAATGCC 2398
697 rcGlyThrTyrAlaThrMetPheLeuIleGlyGluLeuIleGlyGluLys 714
2399 AAGCAACATTTAATCAAGC.....ACATTTAAACGCG 2430
714 eSerIlePheValLysSerHisGlnGlyTyrThrPheLysThrAspGly 730
2431 AACNCATCGGNTTCGGCAATGCTTCA..... 2457
731 AsnThrLysIleAlaGlyLysTyrSerLysGluAspLeuLysIleThrAl 747
2458 .TTTAATCTAAGCAACACGCGCACAA.....AAGCGCAGTC 2494
747 alIeAsnThrGlyGlyArgAlaAlaGluGluValLeuIleAsnGlyAlaL 764
2495 TGACGCTTCGCAACAGCTAAGGCAACGTAAGCATTCGCACTCAAC 2544
764 euGlySerAlaAspAsn...AspAlaAsnIleAlaAsnMetAlaPhe... 778
```

```

2545 GGCAATGTCCTCCAGCCGATAGCAGATATTCATTGTTGAAACACCG 2594
779 .....ThrlieglyaspSerAla.....AsnThrly 787
2595 CTTTACGGGACACTCAGCGGCGACAGAGAACGATTCACATTCAAAAG 2644
787 sThrThr.....IlegluAsnAlaAspIleThrAlaLeu..... 798
2645 ACAGCGATGAGCGCTGCCGCGACGACGCGATGACCAATTTAAACCTT 2694
799 .....AlaProAsnSly.....GlyThrAlaTyrLeu 807
2695 GACAACGCCACCATTAACACTC.....AATTCGCCCTATCCGCA 2732
808 SerSerlyaspValGluIleGluValLysProAsnSerAsnPro..... 822
2733 CGATGCTCAGACGGCGGCAACCGGAGNGTGTACAGACACCGCGCGCC 2782
823 .....ThrPheGluLeuProArg..... 829
2783 GTTCGGCGCGTCCCTATTATCCGTACACCGGCACTTCGCTGATTC 2832
830 .....GluLys 831
2833 GCTTCAACAGCGTGCAGCGTAACGCG.....AAATGAACNGTCA 2873
832 AsnLeuAsnGlnThrLysIleAsnGlyAlaSerThrLysLeuSerGluAr 848
2874 AGGAACATTCGCGCTTATGTCGGAACCTTCGCTACCGAGCAACAAT 2923
848 gGlyPheAlaIleGlyLeuIleAsnGlyAlaIleGlyAlaSerAsnL 865
2924 TGAACCTGGCGGAAGTCCGACAGGNACTACACTTCGCGGCTCAACAT 2973
865 eu..... 865
2974 ACCGGCAACGAAACCCGTAAGCTCGATCATTCAGGTAGTGAAGGAA 3023
866 .....SerAlaGluGlnIleAsnValThrAspAlaSer 876
3024 AGACACAAACCGCTGTCGGAACCTTAATTCACCTCGCAAAACGAC 3073
876 rGlu...LysIleIleAsnThrLysLeuValSerSerLeuAspValGluL 892
3074 ACGTC.....GATCGCGCGCGGTGCGCTTACCAACTC 3105
892 yslLeuValSerValAlaValLysAspAlaGly..... 902
3106 ATCCGCAAAAGACGGCGAGTTCGCTGATATATCCGTCGCAAGAACAGA 3155
903 .....AsnGlyCysGluGlnGlnI 909
3156 GGTTCGCGACAAA.....CTCGGCAACGACAGAG 3184
909 nPheGlyaspLysGlyAsnAsnThrLysValSerValGlyLeuGlnUA 926
3185 CCAAAAAA 3192
926 laGluGln 928
seq_name: SwissProt_40:FRPC_NEIMC
seq_documentation_block:
ID FRPC_NEIMC STANDARD; PRT; 1829 AA.
AC P5127;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Iron-regulated protein frpc.
GN FRPC.
OS Neisseria meningitidis (serogroup C).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=135720;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FAM20 / SEROGROUP C;
RX MEDLINE=94018616; PubMed=8412674;
RA Thompson S.A., Wang L.L., Spallling P.F.;
RT "Cloning and nucleotide sequence of frpc, a second gene from
RT Neisseria meningitidis encoding a protein similar to Rtx
RT cytotoxins."
RL Mol. Microbiol. 9:85-96(1993).
CC -1- FUNCTION: MAY PARTICIPATE IN THE PATHOGENESIS OF MENINGOCOCCAL
CC DISEASE.
CC -1- SUBCELLULAR LOCATION: OUTER-MEMBRANE ASSOCIATED AND SECRETED.
CC -1- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING
CC CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC
CC ACTIVITY.
CC -1- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L06299; AAA99902.1; -
DR InterPro: IPR001343; HemLysn_Ca_Bind.
DR Pfam: PF00353; hemolysinCbind; 9.
DR PRINTS: PR00313; CARNDNGRPT.
DR PROSITE: PS00330; HEMOLYSIN_CALCIUM; 17.
KW Toxin; Calcium; Outer membrane; Repeat.
KW DOMAIN 879 1702 43 X REPEATS, GLY-RICH.
FT REPEAT 879 884 1.
FT REPEAT 888 893 2.
FT REPEAT 897 902 3.
FT REPEAT 1016 1021 4.
FT REPEAT 1025 1030 5.
FT REPEAT 1034 1039 6.
FT REPEAT 1043 1048 7.
FT REPEAT 1052 1057 8.
FT REPEAT 1061 1066 9.
FT REPEAT 1070 1075 10.
FT REPEAT 1079 1084 11.
FT REPEAT 1088 1093 12.
FT REPEAT 1097 1102 13.
FT REPEAT 1216 1221 14.
FT REPEAT 1225 1230 15.
FT REPEAT 1234 1239 16.
FT REPEAT 1243 1248 17.
FT REPEAT 1252 1257 18.
FT REPEAT 1261 1266 19.
FT REPEAT 1270 1275 20.
FT REPEAT 1279 1284 21.
FT REPEAT 1288 1293 22.
FT REPEAT 1297 1302 23.
FT REPEAT 1416 1421 24.
FT REPEAT 1425 1430 25.
FT REPEAT 1434 1439 26.
FT REPEAT 1443 1448 27.
FT REPEAT 1452 1457 28.
FT REPEAT 1461 1466 29.
FT REPEAT 1470 1475 30.
FT REPEAT 1479 1484 31.
FT REPEAT 1488 1493 32.
FT REPEAT 1497 1502 33.
FT REPEAT 1616 1621 34.
FT REPEAT 1625 1630 35.
FT REPEAT 1634 1639 36.
FT REPEAT 1643 1648 37.
FT REPEAT 1652 1657 38.
FT REPEAT 1661 1666 39.
FT REPEAT 1670 1675 40.
FT REPEAT 1679 1684 41.

```





```
1324 .....GTAAACGGCGTGGCAACGACCGCTGTCCAA 1355
1325 LysAspIleLeuArgPheThrAspGlyIleThrAlaAspMetLeuThrPh 1341
1356 AATCGCAAAAGGCAGC...CTGCACGCTTCACCAAGGGAACCAAG 1402
1341 eHrtaFgIuGlyAsnHisLeuLeuLeuLysAlaLys...AspAspSerG 1357
1403 GCTCGATCGACGCTG.....GGCGACGGTACAGTC 1431
1357 LysIleValIleThrValGlnSerTyrPheGlnAsnAspGlySerGlyAlaTyr 1373
1432 ATTTTGATCAGCAGCAGCATTAAGCAAAACAAACAGCTTATAGGA 1481
1374 ArgIleAspGluIleHisPheAspAsnGlyLys.....Va 1385
1482 AATCGCTTGTCAGCGGCGGAGGTACGTCGCACTGAATGCCGTAATC 1531
1385 IleuAspValAlaThrValGluLeuValGlnGlnSerThrAspGlyLys 1402
1532 AGTTCAACCCGCAAACTCTATTTCGCTTCGGCGGGGAGCTTGAT 1581
1402 et.....AspArgLeuTyr...AlaTyrGlnSerGlySer...Thr 1413
1582 TTAAACGGCGCATTCGCTTCGCTTCACCGTATTCAAATACCGATGAAG 1631
1414 LeuAsnGly...GlyLeuGlyAspAspTyrLeuTyrGlyAlaAsp...G 1428
1632 GCGCATGATGNCATTCATATGCCACAACAATCCACCGTTACCATTA 1681
1428 YAspAspLeuLeuAsnGlyAspAlaGlyAsnAspSerIleTyrSerGlyA 1445
1682 CAGGAATGAAGATATTACACACCGGATGGAAGATATATGATGACTT 1731
1445 snGlyAsnAspThrLeuAspGlyGlyGlyGly...AsnAspAlaLeuTyr 1460
1732 AATTACACCAAGAAATTCCTACACCGTGTGGTGGCGAAGAAATAC 1781
1461 GlyTyrAsnGlyAsnAspAlaLeuAsnGlyGlyGly..... 1473
1782 GACCAAAAGCAAGCGGCGCTCAACCTGTTTACCAGCCCGCCAGAG 1831
1474 .....AsnAspHisLeuAsn.....GlyIleUA 1481
1832 ACCGACCCNGCTGCTTCGCGGCAACAATTA..... 1866
1481 spGlyAsnAspThrLeuIleGlyGlyAlaGlyAsnAspTyrLeuGly 1497
1867 ...AAGCGCAATCACCGCAACAAGGCAACTGTTTTCACGGCAG 1913
1498 GlySerGlySerAspThrTyrValPheGlyLysGlyPhe.....GlyG 1512
1914 ACCGACACCGCAGCTTACATCATTTAGCAAGCGGAGGTCAAAATG 1963
1512 nasPThrValTyrAsnTyrAspTyr..... 1520
1964 AAGGATCCCAAGAGAAATCGTGGGCAACGACTGATCAGTCACCGC 2013
1521 ..AlaThrGlyArgLysAspIleLeuArgPheThrAspGlyIleThrAla 1536
2014 .....ACGTTTAAAGCGAAATTTCCATATTTCAGGGCGGCGAGC 2054
1537 AspMetLeuThrPheThrArgGlyGlyAlaHisLeu..... 1549
2055 GGTGATTTCCCGCAATGTGGCAAGTGAAGCGGATTCGATTTGAGCA 2104
1550 .....IleLysAlaLysAspGlySerGlyGlnValThr. 1560
2105 ATCAGCGCCAAGCAGTTTGGTGTCCACCGCATCAAGCATACATC 2154
1561 .....ValGlnSerTyrPhe..... 1565

2155 TGTAACGCTTCGACGTGACNGCTGTGCAAAATTTGTGCAANAANCAAT 2204
1566 .....GlnAsnAspGlySerGlyAlaTyrArgIleAspGluIleHisPh 1580
2205 TACCGAGATTAAGTG.....ATTGCTTCATGACTTAAC..... 2238
1580 eAspAsnGlyLysValLeuAspValAlaThrValLysLysLeuValGlnG 1597
2239 .....ACNGACNTNAGCGCANTGTNAGCTNNCANTNACGNTNNTTNA 2283
1597 InsThrAspGlySerAspArgLeuTyrAlaTyrGlnSerGlyAsnThr 1613
2284 AANTCTCNGCGCTGCN.....NCACTNANAGC..... 2313
1614 LeuAsnGlyGlyLeuGlyAspAspTyrLeuTyrGlyAlaAspGlyAspAs 1630
2314 .....AATCTTAGTCAATGCGCATATACGTTATATACGTCACGACA 2356
1630 PLeuLeuAsnGlyAspAlaGlyAsnAspSerIleTyrSerGlyAsnGlyA 1647
2357 ACGCCACCCAAAC.....GCGAACCTTAGCTCTGTGGCAATGCC 2397
1647 snAspThrLeuAsnGlyGlyGlyGlyAsnAspAlaLeuTyrGly..... 1661
2398 CAAGCAACATTATATCAAGCCACATTAAACGGCAACNATCGGNTTCGG 2447
1662 .....TyrAsnGlyAsnAspValLeuAsnG 1670
2448 CAATGCTTCATTATATCAATCAACCAACCGCCGCAAAAGCGAGCTCTA 2497
1670 YGlyGlyGlyAsnAspHisLeuAsn.....GlyGlyAspGlyAsnAspT 1685
2498 CGCTTCGCGCAACGCTTAAGCAACGTAAGCCATTCCGCACTCAAGGC 2547
1685 hrLeuIleGlyGlyAla.....GlyAsnAspTyrLeuGlyGly 1697
2548 AATGTCCTCTACCGATGAAGCGATATTC..... 2577
1698 Gly...SerGlySerAspThrTyrValPheGlyLysGlyPheGlyGlnAs 1713
2578 .....CATTTGAANAACAGCGCTTACCAGCAACCTCA 2611
1713 rPhrValTyrAsnTyrHisValAspLysAsnSerAspThrMetHisPhe 1730
2612 GCGCAGCAAGGANNACGATTAAC...TTAAACACAGCGCAATGAGC 2658
1730 YsGlyPheLysAlaAlaAspValHisPheIleTyrSerGlySerAspLeu 1746
2659 CTGCGCTCAGGCAAGCAATTAAGCAATTTAAACCTT..... 2694
1747 ValLeuSerAlaSerGlyGlnAspAsnValArgIleSerGlyPhePheTyr 1763
2695 .....GACAAGCCACCATTA 2710
1763 rGlyGlnAsnHisArgValAspThrPheValPheAspAspAlaAlaIle 1780
2711 CACTCAATTCGCGCTATCGCAGATGCTGACGCG..... 2745
1780 eAsnProAspPheAlaLysTyrIleAsnAlaGlyAsnAsnLeuValGln 1796
2746 .....GCGCAACCGCGCAGNGTGCAGA 2768
1797 SerMetSerValPheGlySerAsnThrAlaIleThrGlyLysValAs 1813
2769 CACGCGCGCGCGCGCTGCGCGCTTCCTATTATCCGTTACAGC 2814
1813 pAlaAsnThrGlnSerValGlnGlnProLeuLeu...ValThrPro 1827

seq_name: SwissProt_40:OMP_RICRI
seq_documentation_block:
ID OMP_RICRI STANDARD; PRT; 1654 AA.
AC Q53047;
```

DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)  
 DE (surface protein antigen) (cell surface antigen 5) (Scas5) (rOmpB)  
 DE (rOmp B) [contains: 120 kDa surface-exposed protein (Surface protein  
 antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].  
 GN OMPB.  
 OS Rickettsia rickettsii.  
 OC Bacteria: Proteobacteria: alpha subdivision: Rickettsiales;  
 OC Rickettsiaceae: Rickettsiense; Rickettsia.  
 OX NCBI\_TaxID=783;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=R;  
 RX MEDLINE=92167802; PubMed=1724278;  
 RA Gilmore R.D. Jr., Cleplak W. Jr., Policastro P.F., Hackstadt T.;  
 RT "The 120 kDa outer membrane protein (rOmp B) of Rickettsia  
 rickettsii is encoded by an unusually long open reading frame;  
 RT evidence for protein processing from a large precursor";  
 RL Mol. Microbiol. 5:2361-2370(1991).  
 RN [2]  
 RP SEQUENCE OF 279-1654 FROM N.A.  
 RC STRAIN=R;  
 RX MEDLINE=90136087; PubMed=2515418;  
 RA Gilmore R.D. Jr., Joste N., McDonald G.A.;  
 RT "Cloning, expression and sequence analysis of the gene encoding the  
 RT 120 kD surface-exposed protein of Rickettsia rickettsii";  
 RL Mol. Microbiol. 3:1579-1586(1989).  
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
 CC LAYER WITH HEXAGONAL SYMMETRY.  
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: X16353; CA34403.1;  
 DR InterPro: IPR003858; rOmpA\_rOmpB.  
 DR Pfam: PF02708; rOmpA\_rOmpB; 1.  
 KW Antigen; S-layer; Cell wall.  
 FT CHAIN 1 1333 120 KDA SURFACE-EXPOSED PROTEIN.  
 FT CHAIN 1334 1654 32 KDA BETA PEPTIDE.  
 FT DOMAIN 1181 1188 POLY-TMR.  
 SO SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;

alignment\_scores:  
 Quality: 198.50 Length: 1660  
 Ratio: 0.289 Gaps: 69  
 Percent Similarity: 41.446 Percent Identity: 17.831

alignment\_block:  
 US-09-303-518d-651 x OMPB\_RICRI ..  
 Align seg 1/1 to: OMPB\_RICRI from: 1 to: 1654

217 GTTACACACAAAAAGGAGTGTGGCGCAATCATGACAAAAGCCC 266  
 |||:||||| |||:||||| |||:|||||  
 185 VALPheAsn.....LeuAlaAsnProThrInGlnLysAlaDr 197  
 267 GATGATTTGATTTTCTGTGTGTCCGCGTAAGGCGCTGGCGGCGATTGGTG 316  
 |||:||||| |||:||||| |||:|||||  
 197 Oeuille.....LeuG 201  
 317 GCGATCAATATATGTGTGAGCGTGACATTAAGCGGCGCTATATAACAGCTT 366

201 |||:||||| |||:||||| |||:|||||  
 201 LysAspSmlaValIleAlaAsnGlyValAsnGlyThrLeuAsnValThr 217  
 367 GATTTGGTGGCGGAAGGAAGNAATCCGATCGACACCGTTTCTTACCA 416  
 |||:||||| |||:||||| |||:|||||  
 218 AsnGlyPheIleGlnValSerAsnLys.....SerPheAl 229  
 417 AATTGTGAAGAAATATATTAATTAACCTTGACAAATTCACACCTTAAACG 466  
 |||:||||| |||:||||| |||:|||||  
 229 aThrValLysAlaIleAsnIleAlaAspGlyGlnGlyIleIlePheAsn. 245  
 467 GCGATTANCAATATGCGCGGTTTGCAATTAATTGTTCACAGATGACAAAGCT 516  
 |||:||||| |||:||||| |||:|||||  
 246 .....ThrAspAlaAsn 250  
 517 GTCGAAATAGCAGATGACATGAGGGGAATACCTTATTCGATTAAGAAA 566  
 |||:||||| |||:||||| |||:|||||  
 251 AlaAsnThrLeuAsnLeuGlnAlaGlyThr..... 261  
 567 ATATCCGAGCGTGTCCGATCGCTCAGACACCATTAATTTGGCGTTATG 616  
 261 ..... 261  
 617 ATGATGACAAACGCGGATTTATCTTCTCGCGC.....GCA 654  
 262 .....ThrIleAsnPheThrGlyThrAspGlyThrGly 272  
 655 TGGTTAATTTGGCGGCAATACATATGCGAGGTTGGGGAATTAATGCGCT 704  
 |||:||||| |||:||||| |||:|||||  
 273 ArgLeuValLeuLeuSerIleHisAlaIleAlaThrAsnPheAsnIleThr 289  
 705 AATTGTGTGAGCGCGGATGTGCGCGCATGCGCAAGCATATGAGCCCTATGC 754  
 |||:||||| |||:||||| |||:|||||  
 289 GylSerLeuGlyGlyAsnLeuLysGlyValIleGlnPheAsnThrValA 306  
 755 CGATTGACGAGTGGCGGCGAGCGAGCGGTGCCAATGTTATTATTATGAC 804  
 |||:||||| |||:||||| |||:|||||  
 306 lAlaLysPgly.....Gln 310  
 805 AAACAAACAAATTAATGCTGCTCAACGAGTTTACAAACGCGCTACCC 854  
 |||:||||| |||:||||| |||:|||||  
 311 LeuThrAlaAsnAlaGlyAlaAlaAsnAlaValIleGlyThr..... 324  
 855 TTATTCCGCGGAGGAAACGGTTTCCAGCTGATACGCAAGATTGTTCT 904  
 324 ..... 324  
 905 ACGATGACATTACAGAGCGCATACATACCGTCTTTTGTGAACCGCGC 954  
 324 ..... 324  
 955 AGTACGAGCATTTTCTTCTTACATCCACACACAGCGTACGGGT..... 999  
 325 .....AsnAsnGlyAlaGlyArgAl 331  
 1000 .....ACGTAACAGAAACACAGAAAGGTTTCCCAATCCAAAGC 1039  
 |||:||||| |||:||||| |||:|||||  
 331 aAlaGlyPheValValSerValAspAsnGlyValAlaIleThrIleAspG 348  
 1040 TTAAGTACAGACAGTCCGACTGTTTGACGAATCTTGAATGAACATGAT 1089  
 |||:||||| |||:||||| |||:|||||  
 348 LysGlnValValAlaLysAspMetValIleGlnSerAlaAsnAlaIleThrGly 364  
 1090 AAGAACACAGTT.....TACCGGACAGGGGCTGTATCA 1124  
 |||:||||| |||:||||| |||:|||||  
 365 GlnValAsnPheArgHisIleValAspValGlyValAspGlyThrAla 381  
 1125 GTACCGTCCA.....AGTTAAACAAGCGGAACCTT 1159  
 |||:||||| |||:||||| |||:|||||  
 381 aPheLysThrAlaLaserLysValThrIleThrGlnAspSerAsnPheG 398  
 1160 CTTTATTCGATTACGGCAAC.....GCCAAA 1185  
 |||:||||| |||:||||| |||:|||||

398 LysThrAspPheGlyAsnLeuAlaIleLysValProAsnAla 414  
1186 CTCATCTTATCAACACATC.....AACCAAGC... 1215  
415 IleThrLeuThrGlyAsnPheThrGlyAspAlaSerAsnProGlyAsnTh 431  
1216 .GGGGGGCTTTGATTTGAAAGTATTTAGCTCGCTCGCTGAAACA 1264  
431 rAlaGlyValIleThrPheAspAlaAsnGlyThrLeuGluSerAlaSer 448  
1265 ACGAAGCTGGCAAGCGCGGCTCATATCAGTGAAGACATGCGTT 1314  
448 laasp.....AlaasnValAlaValThrAsnAsnIleThrAla 460  
1315 ACTTGAAGTAACAGCGCTGGCAAGACCGCTGCCAAATCGGCA 1364  
461 Ile..GluAlaSerGlyAlaGlyValValGlnLeuSer..... 472  
1365 AGGACGCTGCACGTTTCAGCCAAAGGGAACCAAGGCTCGATC...A 1411  
473 .GlyThrIleAlaIleGluLeuArgLeuGlyAsnAlaGlySerIlePhe 489  
1412 CGGTGGCGGACGCTACATCTTTGGATCGACGACGACGATAAAGC 1461  
489 yslauAlaAspGlyThrValIle.....AsnGly 498  
1462 AAAAAACAAGCCTTAGTAATCGGCTTGATCAGCGGAGGCTAGCGT 1511  
499 LysValaAsnGlnThrAlaLeuValGlyGlyAlaLeuAlaIleGlyThr 515  
1512 GCAACTGAATGCCGATTAATCAGTTCACCCGAC..... 1545  
515 eThrLeuAspGlySerAlaThrIleThrGlyAspIleGlyAsnAlaGly 532  
1545 ..... 1545  
532 lYAlaAlaLeuGlnArgIleThrLeuAlaAsnAspAlaLysLysThr 548  
1546 .....AACTCTATTTCGGCTTTCGGCGGCGGACGCTTGA 1580  
549 LeuThrLeuGlyGlyAlaAsnIleIleGlyAlaGlyGlyThrIleAs 565  
1581 TTTA.....AAGGGCATTCGCTTTCGTCACCGTATTCAAAT... 1620  
565 pLeuGlnAlaAsnGlyGlyThrIleLysLeuThrSerThrGlnAsnAsn 582  
1621 .....ACGATGAAGGGGCGATGAT 1641  
582 leValValAspPheAspLeuAlaIleAlaThrAspGlnThrGlyValAl 598  
1642 GNCNATCATATGCGACACAACATCACCGCTTACCATTTACAGGATGA 1691  
599 AspAlaSerSerLeuThrAsnAlaGlnThrLeuThrIleAsnGlyLys 615  
1692 AAGTATTACACAACGAGTGTAGATATCATATGACTTAATTC... 1737  
615 eGlyThrIleGlyAlaAsnAsnLysThrLeuGlyGlnPheAsnIleGly 632  
1738 ..ACGAAGAATATGCTACAAAGGTTGTTGGGAGAGAAGAT... 1779  
632 eSerLysThrValLeuSerAsnGlyAsnValAlaIleAsnGlnLeuVal 648  
1779 ..... 1779  
649 IleGlyAsnAspGlyAlaValGlnPheAlaIleAspThrTyLeuIleTh 665  
1780 ....ACGACCAAGCAAGCGGCGCTCACTGTTTACAGCGCGCGG 1825  
665 rArgThrThrAsnAlaIleGlyGlnGlyLysIleIlePheAsnProVal 682  
1826 CAGAAGACCGACCCNGCTGCTTCGGGAGCAAAATTTAAACGGC... 1872  
682 alAsnAsnGlyThrThr...LeuAlaIleGlyThrAsnLeuGlySerAla 697  
1872 ..... 1872  
698 ThrAsnProLeuAlaGluIleAsnPheGlySerLysGlyValAsnValAs 714  
1873 .....ANATCA 1879  
714 pThrValLeuAsnValGlyGlyGlyValAsnLeuTyralaThrAsnIle 731  
1880 CGCAAAACAAC.....GGCAAACTGTTTCAGCGCGGACGACCGCA 1920  
731 hrThrThrAspAlaAsnValGlySerPheValPheAsnAlaGlyGlyThr 747  
1921 CGCACCGCTTACATCATTTAGAACCGGCTGTCMAAATGAAGAT 1970  
748 AsnIleValSerGlyThrValGlyGly.....GlnGlnGlyAs 760  
1971 CCCACAGAGAAATCGTGTGGACACGACGTGATCACC... 2013  
760 nLysPheAsnThrValAlaLeuGluAsnGlyThrThrValLysPheLeu 777  
2014 .....ACGTTTAA.....GCGGAAATTTCCATAT 2040  
777 LysAsnAlaThrPheAsnGlyAsnThrThrIleAlaIleAsnSerThrLe 793  
2041 CAGGCGGCGGACGCGGTGATTCGCCAATGTTGCCAAATGGAAGCGCA 2090  
794 GlnIleGlyGlyAsnTyThrAlaAspCysValAlaSerAlaAspGlyTh 810  
2091 T..... 2091  
810 rGlyIleValGluPheValAsnThrGlyProIleThrValThrLeuAsn 827  
2091 ..... 2091  
827 ysglnAlaIleProValAsnAlaLeuLysGlnIleThrValSerGlyPro 843  
2092 .....TGNCATTTGACGACATCACCGCCCAAC 2117  
844 GlyAsnValValIleAsnGluIleGlyAsnAlaGlyAsnIleHisGlyAl 860  
2118 AGTTTGTGTCGACCCGATCAAGCCATACATC..... 2154  
860 aValThrAspThrIleAlaPheGluAsnSerSerLeuGlyAlaValAl 877  
2155 .....TGT 2157  
877 heLeuProArgGlyIleProPheAsnAspAlaGlyAsnThrMetProLeu 893  
2158 ACACTTCGCGACGACNGGCTGACAAATGT..... 2190  
894 ThrIleLysSerThrValGlyAsnLysThrAlaLysGlyPheAspValPr 910  
2191 .....GTGCAANAANCATTACGACGATGAAGTGA 2221  
910 oSerValValIleLeuGlyValAspSerValIleAlaAspGlyGlnVal 927  
2222 TTGCTTCATTGACACAGACGACNTNAGCGGACGTCAGTGNCCNCT 2271  
927 leGly.....AspGlnAsnAsnIleValGlyLeuGlyLeu 938  
2272 NACGNTNNTNMAANCNC..... 2289  
939 GlySerAspAsnGlyIleIleValAsnAlaThrThrLeuTyralaGly 955  
2290 .....NCGGCGNTGCGNCCATNAAAGCAATCTTA 2320  
955 eSerThrLeuAsnAsnAsnGlnGlyThrValThrLeuSerGlyValAl 972  
2321 GTGCAAAATGCGGATACGTTAT..... 2343  
972 roAsnThrProGlyThrValTyrcGlyLeuGlyThrGlyIleGlyAlaSer 988

```
2344 .....AC 2345
989 LysPheIleValIleThrPheThrThrAspTyrAsnAsnLeuGlyAsnI 1005
2346 AGTCAGCCCAACGCGCCCAAAAGCGCACTTACCTCGTG.....G 2389
1005 eIleAlaThrAsnAlaThrIleAsnAspIleValThrValThrThrGlyG 1022
2390 GCAATGCCCAACACATTTAATCAAGCCACATTAAGCGCAACACATCG 2439
1022 LylIleAlaIleIleGlyPheAspGlyLysIleThrLeuGlySerValAsn 1038
2440 GNTTCGGGCAATGCTCATTTAAT.....CTAGCAACACAGCGCGC 2480
1039 GlyAsnGlyAsnValAlaArgPheAlaAspGlyIleLeuSerAsnSerThrSe 1055
2481 ACAAAACGCGAGCTGACGCTTCCGACACAGCTAAGCAACAGTAAGCC 2530
1055 rMetIleGlyThr...ThrLysAlaAsnAsnGlyThrValThrTyrLeuG 1071
2531 ATTCCGCACTCAAGCAATGTCCTACCGCATAGGCAATATTCAT 2580
1071 LysAlaIlePheValGlyAsnIleGlyAspSerAspThrProVal..... 1085
2581 TTGAAACACGCGCTTACCGCACTACGCGACGCGACAGANACAGC 2630
1086 ...AlaSerValArgPheThrGlySerAspSerGlyAlaGlyLeuGlnG 1101
2631 A.....TTACACTTAA 2641
1101 ysnIleTyrSerGlnValIleAspPheGlyThrTyrAsnLeuGlyIleV 1118
2642 AAGACAGCAATGAGAGCGTCCGCGTACGACGCGAATTAGCAATTAAC 2691
1118 aIaSerAsnIleIleLeuGlyGlyGlyThrAlaIleAsnGlyLys 1134
2692 CTGGAACACGCGCACCATTAACATCAATTCGCGTACCGACAGATGTC 2741
1135 IleAspLeuValThrAsnThrLeuThrPhe.....AlaSe 1146
2742 AGCGCGCAACCGGACAGNGTGTACAGACAGCGCGCGCGTTCGCC 2791
1146 rGlyThrSerThrThrGlyAsnAsn..... 1154
2792 GTTCCCTATTATCGTTACACGCGCACTCGGTAGATCCGTTTCAAC 2841
1155 .....ThrSerIleGluThr..... 1159
2842 ACGTACGCGTA...AACGCAAAATTGAAC.....NGTCAAG 2876
1160 ThrLeuThrLeuAlaAsnGlyAsnIleGlyHisIleValIleLeuGlnG 1176
2877 AACATTCCGCTTATGTGGAACTCTTGGCTACCGAAGCAACAATTGA 2926
1176 yAla.....G 1178
2927 AGCTGGCGAAAGTTCCGAAGNACTTACACTTGGCGGTCAACAATACC 2976
1178 InValAsnThrThrThrThrThrThrThrIleValAlaGlnAspAsn 1194
2977 GGCACAGAACCGGTAGCCTGATCA...TTGACGGTAGTGAAGG... 3021
1195 AlaAsnAlaAsnPheSerGlyThrGlnThrThrLeuIleGlnGly 1211
3022 ....AAAGCAACAACCGCTG...TCCGAACACCTTAATTCACCCGCG 3064
1211 yAlaArgPheAsnGlyThrLeuGlySerProAsnPheAlaValThrGlyS 1228
3065 AAACGACACACGTGATGCGCGCGGTGCGTTACCAATCATCCGCAAA 3114
1228 eTAsnArgPheValAsn.....TyrSerLeuIleArgAla 1239
3115 GACGGCGAG.....TTCCGCTGCAATAATCCGGT 3143
1240 AlaAsnGlnAspTyrValIleThrArgThrAsnAsnAlaGlnAsnValVa 1256
3144 CAAGAACACAGAGCTTTCGACAAACTCGGCAGAGCAAGCAACCAAAAC 3193
1256 LThrAsnAspIleAlaAsnSerProPheGlyGlyAlaProGlyValAsp. 1272
3194 AGCGGAAAAAGACACCGCCCAAGCCTTGACGCGCTGATGGCGCGG 3243
1273 .....GlnAsnValThrThrPheValAsnAlaThr 1282
3244 CGCGATCGCGCGAAAGACAGAAAGCGTTGCGCAA...CCGCGCGCGCN 3290
1283 AsnThrAlaAlaTyrAsnAsnLeuLeuAlaLysAsnSerAlaAsnSe 1299
3291 GCGACGCGGGAATGTGCGCATTTATGACGCGGAGAGAAAGAAAAAC 3340
1299 rAlaAsnPheValGlyAlaIleValThrAspThrSerAlaAlaIleThr 1316
3341 GGTGTACAGCGCGATTAAGACAGCGCCTTGGCGAAACAGCCGAGGAA 3390
1316 snValGlnLeuAsp.....LeuAlaLysAspIleGlnAlaGln 1328
3391 ACCGCGCGGNTACACCGCTTCCCGCGCGCGCGCGCGCGGGA 3440
1329 .....LeuGlyAsnArgLeuGlyAlaLeuArgTy 1338
3441 TTTCGCGCAACCGCAGCGCCACCGCACTCAACCCCAACCGCAGCGG 3490
1338 rLeuGlyThrProGlnThrAlaGlnMetAlaGlyProGlu..... 1351
3491 ACCTGATNAGCGTTATGCAATAGCGGTTGACTGATTTCCGCGCAG 3540
1352 .....AlaGlyAla 1354
3541 CTCACAGCGTTTTCGCGGTACAGAGCAA...TTGACCGCGCTGTTGC 3587
1355 IleSerAlaAlaValAlaAlaGlyAspGlnAlaIleAspAsnVal..... 1369
3588 CGAAGACCGCGCACGCGCTTGGACAGACNGCATCCGCGNACACCAAC 3637
1370 .....AlaTyrGlyIleThrPalatLysProPheTyrThrAspAlaH 1383
3638 ACTAC...CGTTCGCAAGATTTCCGCGCTACCGCCACCAACACGACCTG 3684
1383 IsGlnSerLysLysGlyGlyLeuAlaGlyTyrLysAlaLysThrThrGly 1399
3685 CGCCAAATCGGTATGCAAGAAAAACCTCGCAGCGGCGCTCGGCATCT 3734
1400 ValValIleGlyLeu..... 1404
3735 GTTTTGCACACCGGACCGCAAAACNCTTGCAGACGCGCATCGCACT 3784
1405 .....AspThrLeuAlaAsnA 1410
3785 CGCGACGCTTTCGCGACGCGCGCTTTCGGCAATACGCGCATCGGAG 3834
1410 sPanLeuMetIleGlyAlaAlaIle.....GlyIleThrLys 1422
3835 TTGCAATCGGCATTCAGCAGCGGCGGCTTTAGCAGCGCANTCTNTC 3884
1423 ThrAspIle.....LysHisGlnAspTyrLysGlyLysPylsThr 1436
3885 AGACGCGATGAGAGCAAAATCCGCGCGCGCTGCTGATTCAGGATTC 3934
1436 rAspValAsnGlyPheSerPheSerLeu.....TyrGlyAlaG 1449
3935 AGCGACGA.....TACCGCGCGGTTTCGCGGATTCGCGCATC 3972
1449 GlnGlnLeuValLysAsnPhePheAlaGlnGlySerAlaIlePheSerLeu 1465
3973 GAACGCTACATCGGC...GCAACGCGCTATTTGTC..... 4005
```

```

1466 AsnGlnValIValAsnLysSerGlnArgTyrPhePheAspAlaAsnGlyAs 1482
4006 .....CAAAAGCGGATTACCGTACGAAAGCGTCAATATACGCCA 4045
1482 mterLysGlnIleAlaIaGlyHisTyrAspAsnMetThr..... 1496
4046 CCCCCGCTTCGCTTCAACCGTACCGGCGGCGATTAAAGCAGATTAT 4095
1497 .....PheGlyGlyAsnLeuThrValGlyTyr 1505
4096 TCATTCAAAACGGCGCAACACATNTCATCNCCTTATTNAGCCTGTC 4145
1506 AspTyrAsnAlaMetGlnValLeuValThrProMetAlaGlyLeuSe 1522
4146 CATATACCGATCGCGCTCGGCAAGTCCGAACACGCGTCAATACCGCG 4195
1522 TrrTLeuLysSerSerAspGluAsnTrrLysGlnThrGlyThrValA 1539
4196 TATTGGCTCAGATTTCGGCAAA...ACCCGCAATGGCGGCGCTA 4242
1539 LAsnLysGlnValAsnSerLysPheSerAspArgThrAspLeuIleVal 1555
4243 AACCGCGAAATCAAGGTTTCACGCTGTC 4272
1556 GlyAlaLysValAlaGlySerThrMetAsn 1565

seq_name: SwissProt_40:N214_HUMAN

seq_documentation_block:
ID N214_HUMAN STANDARD; PRT; 2090 AA.
AC P36558;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear pore complex protein Nup214 (Nucleoporin Nup214) (214 kDa
  nucleoporin) (CAN protein).
GN NUP214 OR CAN OR CAN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=92195315; Pubmed=1549122;
RA von Lindern M., Fornerod M., Van Baal S., Jaegle M., De Wit T.,
  Buijs A., Grosveld G.;
RT "The translocation (6;9), associated with a specific subtype of acute
  myeloid leukemia, results in the fusion of two genes, dek and can,
  and the expression of a chimeric, leukemia-specific dek-can mRNA.";
RL Mol. Cell. Biol. 12:1687-1697(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=94151361; Pubmed=8108440;
RA Kriemer D., Wozniak R.W., Blobel G., Radu A.;
RT "The human CAN protein, a putative oncogene product associated with
  myeloid leukemogenesis, is a nuclear pore complex protein that faces
  the cytoplasm.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:1519-1523(1994).
CC -1- FUNCTION: MAY SERVE AS A DOCKING SITE IN THE RECEPTOR-MEDIATED
  IMPORT OF SUBSTRATES ACROSS THE NUCLEAR PORE COMPLEX.
CC -1- SUBUNIT: HOMODIMER. INTERACTS WITH DDIX19.
CC -1- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX, CYTOPLASMIC FILAMENTS,
  KIDNEY, BRAIN AND TESTIS, BUT HARDLY IN ALL OTHER TISSUES OR IN
  WHOLE EMBRYOS DURING DEVELOPMENT.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THYMUS, SPLEEN, BONE MARROW,
  KIDNEY, BRAIN AND TESTIS, BUT HARDLY IN ALL OTHER TISSUES OR IN
  WHOLE EMBRYOS DURING DEVELOPMENT.
CC -1- DOMAIN: CONTAINS MANY X-X-F-G REPEATS.
CC -1- PTM: PROBABLY GLYCOSYLATED AS IT REACTS WITH WHEAT GERM AGGLUTININ
  (WGA).
CC -1- DISEASE: IMPLICATED IN A SUBSET OF ACUTE MYELOID LEUKEMIA (ACUTE
  NONLYMPHOBLASTIC LEUKEMIA) (AML) CARRYING A CHROMOSOMAL
  TRANSLOCATION t(6;9)(p23;q34) THAT RESULTS IN THE FORMATION OF A
  DEK-CAN FUSION GENE.

```

```

CC -1- DISEASE: IN A CASE OF ACUTE UNDIFFERENTIATED LEUKEMIA (AML) A
  TRANSLOCATION RESULTS IN THE FORMATION OF A SET-CAN FUSION GENE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X64228; CAA4535.1; -.
DR PIR: S26058; S26058.
DR MIM: 114350; -.
DR InterPro: IPR004325; Nucleoporin_FG.
DR InterPro: IPR001680; WD40.
DR Pfam: PF03093; Nucleoporin_FG; 17.
DR SMART: SM00320; WD40; 1.
KW Nuclear protein; Transport; Proto-oncogene; Chromosomal translocation;
  Repeat; Glycoprotein.
FT DOMAIN 481 2076 11 x 5 AA APPROXIMATE REPEATS.
FT DOMAIN 1409 2084 18 x 4 AA APPROXIMATE REPEATS.
FT DOMAIN 1427 2085 11 x 3 AA APPROXIMATE REPEATS.
FT DOMAIN 1213 2090 PRO/SER/THR-RICH.
FT DOMAIN 740 768 LEUCINE-ZIPPER 1.
FT DOMAIN 861 882 LEUCINE-ZIPPER 2.
FT SITE 812 813 BREAKPOINT.
SQ SEQUENCE 2090 AA; 213766 MW; 6DBE67FDB857F8F CRC64;

```

```

alignment_scores:
  Quality: 197.00      Length: 1281
  Ratio: 0.335        Gaps: 63
  Percent Similarity: 45.902  Percent Identity: 20.921

```

alignment\_block:

US-09-303-518d-651 x N214\_HUMAN ..

Align seg 1/1 to: N214\_HUMAN from: 1 to: 2090

```

1082 AACTGATTAAGAAACACGTTACCGCGCGGCGTGTATATCATACGCT 1131
      :::::::::::::::::::: ||| :::: |||:::
866 GtullelleasnglnGlnArgLysArgLeuAsnHisLeuValasPserLe 882
1132 CCAAGGTTAAACAACGCGTAA.....ACCTTCTTTATCATTA 1172
      ||| :::: |||
882 uGlnGlnLeuArgLeuTrrLysGlnThrSerLeuTrrPserLeuSerLeu 899
1173 CGGCAACGGCAACATCATCTTATCAACAACATCAACCAAGCGCGGCG 1222
899 lavalProserGlnSerSerIleHisSerPheAspSerLeuGlnSer 915
1223 GTTGTG.....ATTGAGAGTGATTTTACGGTCT..... 1252
      :::: |||
916 LeuLysAsnAlaLeuLeuLysTrrThrIleGlnSerHisThrLysSerLe 932
1253 .....CGCTGAAACCAACGAACGAGCGGCGCGG 1286
932 uProLysValAlProAlaLysLeuSerProMetLysGlnAlaGlnLeuArg 949
1287 CGTTCATATCAGTGAAGACAGTACCGTTACTTGAAAGTAAAGCGGCTG 1336
      |||::: :::: |||
949 snPheLeuAlaLysArgLysThrProProVal..... 959
1337 CAAACGACCGCGCTGTCCAAATCGGCAAGGACCGCTGCACGTTTC.... 1381
      :::: |||
960 ArgSerThrAla.....ProAlaSerLeuSerArgSerAlaPheLeuSe 974
1381 ..... 1381
974 rGlnArgTrrTrrGlnAspLeuAspGluValSerSerThrSerValAs 991
1382 .....AAGCAAGGCGGCAAAACCAAGGCTCGATCAGCGGTG 1417

```

```
991  ergInserLeuGIuserGIuAspAlaIrgThrSerCysLysAspArgIu 1007
      :::::||||:
1418  GCACGCGTACAGTCATTTGGATCGACGACGACGATAAAGCAAAAAA 1467
      ||:::|||||:::
1008  AlaValAlaIglInAla.....ProArgH 1015
      :::::
1468  CAAGCCTTAGTAATCGGCTGN..... 1492
      : ||:::||||:
1015  sAlaProValAlaIrgThrProSerIleGIInProSerLeuLeuProHIsA 1032
      : ||:::||||:
1493  .....TCACGCGACGGGCTA 1507
      ||:::|||||
1032  laAlaProPheAlaLysSerHisLeuValHisGlySerSerProGIuVal 1048
      ||:::|||||
1508  CGGTGCACTGATCGCGATATCACTTCACCCCGCAAACTGATTTC 1557
      |||
1049  .....MetGIyThrSerValAlaThrSerAlaSerLysIleI 1061
      |||
1558  GGCTTTCGGGCGGACGTTGGATTAAACGGGCAATTCGTTGTTCA 1607
      :::::|||||
1061  eProGIInGIyAlaAspSer.....ThrMetLeuAlaThrLysThrV 1075
      |||
1608  CGGTATTCAAATACCGATGAAGGGCGATGTGNCNATTCATATGCCA 1657
      ::|
1075  alLysHisGIyAlaProSerProSerHis.....ProIleSerAlaPro 1089
      |||
1658  CAACACATCCACCGTTACCATTAACGGAATGAAGTATTACACACCG 1707
      |||||
1090  GIInGIInLeuAlaAlaAlaLeuAlaIrgGIInMetAlaSerGIInAlaPr 1106
      |||||
1708  ACTGTAGATATATCAATAGACTTATTAACAGAAAGTAATTCCTACAA 1757
      |||||
1106  oAlaValAsn...ThrLeuThrGIuserThrLeuLysAsnValProGIInV 1122
      |||||
1758  CGGTGGTGTGGCGAGAAAGATACGACCAAAAGACGGCGGCTCAACC 1807
      ::|||:::
1122  alValAsnValGIInGIInLeuLysAsnAsnProAlaThrProSerThrAla 1138
      |||||
1808  TTGTTTACCAGCCCGCGAGAAAGCCGACCCGCTGTTTCGGGCGGA 1857
      ::|||::|
1139  MetGIySerSerValProIySerThrAlaLysThrProHisProValIe 1155
      |||||
1858  ACAAAATTTAAGCGCACATCAACGCAAAACAAACGCTGTTTTCAG 1907
      ::|||:::
1155  uThrProValAlaAlaAsnGIInAlaLysGIInGIySerLeuIleAsnSerL 1172
      |||||
1908  CGGACAGACGACACCGCGCTACATCATTTAGGAAGCGGTGTCA 1957
      |||||
1172  euLysProSerGIyProThrPro.....AlaSerGIyGIIn 1183
      |||||
1958  AATGGAAAGTATCCACAAAGAAATCGTGTGGACACAGCATGTGATC 2007
      |||
1184  .....LeuSerSerGIyAspLysAlaSerGI 1192
      |||||
2008  NACCCGACGTTTAAAGCGGAAATTTCCATATTCAGGGCGGCGACGGCT 2057
      |||||
1192  yThrAlaLysIleGIInThrAlaValThr.SerThrProSerAlaSerGIy 1208
      |||||
2058  GATTTCGCCGCAATGTGGCAAAAGTGGAAAGCGATTGNCATTTGAGCAATC 2107
      ::|||
1209  GIInPhe.....Se 1211
      |||||
2108  ACGCCCAAGCACTTTTGGTGTGACCGCATCAAGCCATACATCTGT 2157
      |||||
1211  rLysProPheSerPhe.....SerProSerGIyThrGIyP 1223
      |||||
2158  ACAAGTTTGGAGTGCACNCGTGTGACAAATTTGTGTGAAANAANCATTAC 2207
      ::|||
1223  HeAsnPheGIyIle..... 1227
      |||||
2208  CGACGATAAAGTATGCTTCATTGACTAAGACNACNTNAGCGGCANTG 2257
      |||
```







1082 yAsnAspThrLeuIleGlyAlaGlyAsnAspTyrLeuGluGly 1099  
779 GCGGTGCGCAATGTTATTATGACAAACAAACAAATTAATGCGTCTC 828  
||||| : : : : :  
1099 ergLySerAspThrTyrValPheGlyLysGlyPheGlyGln 1112  
829 AACGAGATTTCACAAACCGGCTACCTTAT 1113  
: : : : :  
1113 : : : : :  
870 AACGCTTCCAGCTGATACGCAAGATGCTTACAGTACAT 915  
: : : : :  
1125 sAsp 1125  
: : : : :  
1136 lAspMetLeuThrPheThrArgGluGlyAsnHisLeuIleLeuAla 1152  
949 CCGCGCATACGCGACATTTTCC 992  
: : : : :  
1153 LysAspGlySerGlyGlnValThrValGlnTyrTyrPheGlnAsnAspGly 1169  
993 TACGGGTACGGGTAAACGAAACCAAGAAAGTTCACATCCAAAG...C 1039  
: : : : :  
1169 ySerGlyAlaTyrArgLysAspGlyLeuHisPheAspGlnGlyVal 1186  
1040 TTAAGTACAGACAGTCCGA...CTGTTGACGAATCTTGAATGAAC 1086  
||| : : : : :  
1186 euAspValAlaThrValLysGluLeuValGlnGlnSerThrAspGlySer 1202  
1087 GATTAAGAACACAGTTACGCGGCGGTGTTAATCAGTACGCGCCAG 1136  
: : : : :  
1203 AspArgLeuTyrAlaTyrGlnSerGlyAsnThr 1213  
1137 GTTAAC 1153  
: : : : :  
1214 LeuAsnGlyGlyLeuGlyAspAspTyrLeuTyrGlyAlaAspGlyAsp 1230  
1154 ACCTTCTTTTATGATTACGACGCAACATCATCTTATTCACAAAC 1203  
: : : : :  
1230 sPheLeuAsnGlyAspAlaGlyAsnAspSerIleTyrSerGlyAsnGly 1246  
1203 : : : : :  
1247 AsnAspThrLeuAspGlyGlyGluGlyAsnAspAlaLeuTyrGlyTyrAs 1263  
1204 : : : : :  
1263 nGlyAsnAspAlaLeuAsnGlyGlyGluGlyAsnAspHisLeuAsnGlyG 1280  
1240 : : : : :  
1280 lAspGlyAsnAspThrLeuIleGlyGlyAlaGlyAsnAspTyrLeuGln 1296  
1279 GGGCG 1304  
: : : : :  
1297 GlyGlySerGlySerAspThrTyrValPheGlyLysGlyPheGlyGlnAs 1313  
1305 CAGTACCGTACTTGAA 1323  
: : : : :  
1313 pAlaValTyrAsnTyrAspTyrAlaThrGlyAlaGlyAsnAspIleLeuArg 1330  
1324 : : : : :  
1330 heThrAspGlyIleThrAlaAspMetLeuThrPheThrArgGluGlyAsn 1346  
1351 : : : : :  
1347 HisLeuLeuIleLysAlaLysAspGlySerGlyGlnValThrValGlnSe 1363  
1386 CAAGGGGAAACCAAGGCTGATGACGCTGGGCGACGATACAGCATTT 1435  
: : : : :  
1363 rTyrPheGlnAsnAspGlySer 1375  
1436 TGGATACGACGACGACGATTAAGCAAAACAAAGCCTTAGTGAATC 1485  
: : : : :  
1375 lAspGluIleHisPheAspAsnGlyLys 1386  
1486 GCGTGTGTCAGCGGCGAGCGGTACGCTGCAACTGAATCCGATTAACGTT 1535  
: : : : :  
1387 AspValAlaThrValLysGluLeuValGlnGlnSerThrAspGlySer 1402  
1536 CAACCCGCAAACTGATTTGCGCTTTCGCGGCGGAGCTTGGATTAA 1585  
: : : : :  
1403 : : : : :  
1586 ACCGGATTCGCTTTCGTCACCGGATTCAAATTCAGTAAAGGCGG 1635  
1415 smGly 1429  
: : : : :  
1430 AspLeuLeuAsnGlyAspAlaGlyAsnAspSerIleTyrSerGlyAsnGly 1446  
1686 GAATGAAGATTTACACAAACCGGCTGTAAGATATCATATGACTTAAT 1735  
1446 yAsnAspThrLeuAsnGlyGlyGlyGly...AsnAspAlaLeuTyrGlyT 1462  
1736 ACAGCAAGAAATGCTGCAACGCTGCTGCGGAGAAAGATACGACC 1785  
: : : : :  
1462 yTrAsnGlyAsnAspAlaLeuAsnGlyGlyGly 1473  
1786 AAACGACGCGGCGCTCAACCTGTTTACACGCGCGGCAAGACCG 1835  
: : : : :  
1474 : : : : :  
1836 CACCGACGCTTTCGCGGCAACAAATTA 1867  
1482 yAsnAspThrLeuIleGlyGlyAlaGlyAsnAspTyrLeuGluGlyLys 1499  
1868 ACGGCAACATCAGCAACAAACGCAACCTGTTTACAGCGGACAGCG 1917  
: : : : :  
1499 ergLySerAspThrTyrValPheGlyLysGlyPhe 1513  
1918 ACACCGACCGCTCAACATCTTAGAGAGCGGCTGTCAAAAATGGAAG 1967  
: : : : :  
1514 AlAlaValTyrAsnTyrAspTyr 1521  
1968 TATCCCAAGAGAAATCGTGTGGGACAGCATGATCCAGC 2013  
: : : : :  
1521 aThrGlyArgLysAspIleLeuArgPheThrAspGlyIleThrAlaAspM 1538  
2014 : : : : :  
1538 eLeuThrPheThrArgGluGlyAsnHisLeu 1549  
2059 ATTTCCCGCAATGTTGCCAAAGTGAAGGATGATGACATCA 2108  
1550 : : : : :  
2109 CGCCCAAGCAGATTTTGTGTGCGACCGCATCAAGCCATCAATCTGTA 2158  
1561 : : : : :  
2159 CAGCTTGGAGTGCAGCGCTGACAAATGTGTGCAANAANCAATAC 2208  
1566 : : : : :  
2209 GACGATTAAGTATGCTTCATTAAGT 2237  
1582 AsnGlyLysValLeuAspValAlaThrValLysGluLeuValGlnGlnSe 1598  
2238 GACGACATNAGCGGCAANTGATGATGATGATGATGATGATGATGATGAT 2287  
: : : : :  
1598 rThrAspGlySerAspArgLeuTyrAlaTyrGlnSerGlySerThrLeuA 1615



60 SerLysAsnAsnThrSerAsnGlySerSerHis..... 71  
1119 TAATCAGTACCGTCCAAAGGTTAAACAAACGCGTGAACAACTTTCTTTATCG 1168  
72 ..... SerAsnAsnAsnThrSerSerIleIleA 81  
1169 ATTACGGCAACGGCAACTCATCTTATATCAACACATCAACCAACGGCCG 1218  
81 lGluAlaAlaAlaLysPheLeuLeuLysAsnGlyLeuAsn..... 94  
1219 GCGGGTTTATTTTGAAGTGATTTTACGGTCTCGCTGAACAAACAGA 1268  
94 ..... 94  
1269 AACGTGCAAGCGCGCGGCTTATATCATGAGAACACTACCGTTA... 1315  
95 ..... GlySerSerThrSerThrProProLeuPro 105  
1316 ..... CTGGAAGTAGTAACGGCGTGCCAAACGACGCGC 1347  
106 ProProLeuProAlaAsnLeuSerArgThrThrProThrThrThr 122  
1348 CTGTCGCAAAATCGGCAAGGACAGCTGCAGC..... TTCAGCCAAAGC 1391  
122 rThrProSerSerSerSerSerThrAlaSerAsnGlyPheLeuProHisA 139  
1392 GGAACCAACGAGCTCGATCAGCG..... TGG 1417  
139 lAluSerProLysSerSerSerIleMetAlaAlaSerAlaAlaValAla 155  
1418 GCGAGCTACAGTCATTTTGGATCAGCAGCAGCATTAAGGCAGCAAAAA 1467  
156 AlaserAlaValGlyAlaThrAlaSerLysProThrIleAspValLeuG 172  
1468 CAAACCTTAGTGAATCGGCTTGNTCAGCGGAGGGGTACGCTGCAACT 1517  
172 yGlyValLeuAspTyrSerSerLeuGlyGlyAlaAlaThr..... 185  
1518 GAATGCCGATTAATCAGTTCAACCCGACAAACTATTTGCGCTT... 1563  
186 ..... GlySerLeuProThrThrAlaValAlaAlaAlaAla 197  
1564 ..... CGCGCGGAGCGTTTGA 1580  
198 AlaglyThrAlaLysIleGlyLysGlySerAsnSerGlyGlySerPheAs 214  
1581 TTAAACGGCGATTCGCTTCTGCTCCAC..... 1608  
214 pMetGlyArgThrProIleSerThrHisGlyAsnAsnSerTrpGlyLys 231  
1609 ..... CGTATTCAAATAACGATGAGGCGGATGATGNCNAT... 1647  
231 yGlyGlyArgLeuGlnPhePheLysAspGlyLysPheIleLeuGluLeu 247  
1647 ..... 1647  
248 AlaArgSerLysAspGlyAspLysSerGlyTrpValSerValThrArgLys 264  
1648 ..... CATATGCCACAACAACATCCACCGTTTACCATTTACAG 1684  
264 sThrPheArgProProSerAlaAlaThrSerAlaThrValThrProThrS 281  
1685 GGAATGAAGAATTTACACACGAGTGGTAAGAATATCAATAGACTTAAT 1734  
281 eAlaAlaValThrThrAlaLysTrpLysAsnGluAsnSerThrSerLeuSer 297  
1735 TACAGCAA... AGAATTTGCTTACCAACGGTTGTTGGCGAGAAAGATA 1780  
298 PheSerAspAspAsnSerSerIleGlnSerSerProTrpGlnArg... 312  
1781 CGACAAACAAAGACGGCGGCTCAACCTTTTACCAGCCCGCGGAGAA 1830  
313 AspGln..... ProTrpLysGlnSerA 320  
1831 GAACCGACCCNCGCTGCTTCCGCGGAGCAAAATTAAACGCCAA... 1874  
320 rGPro..... ArgArgGlyIleSerLysGlyLeuSerLeu 331  
1875 ...CATACGCAACAAACGCAAACTGTTTTCAGCGGACAGCCAGAC 1921  
332 PhePheHisArgProArgAsnSerThrLeuGlyArgAlaAlaLeuArgTh 348  
1922 CGACGCCCTACATCATCTTTAGGAGAGCGGGTGTCAAAAATGGAAGTATC 1971  
348 rAlaAla..... ArgLysArg..... ArgArgProHisGluP 359  
1972 CCACAAAGAGAAATCGTGTGGACAAACGACGTGATCCACCGACTTAA 2021  
359 rLeuThrThrSerGluAspGlnGlnProIlePheAlaThrAlaLeu 375  
2022 AGCGGAAAT..... TTCATTTTCAAGCGCGGACGCGGTGA 2059  
375 sAlaGluAsnGlyAspAspThrLeuLysAlaGluAlaAlaGluAlaValG 392  
2060 TTTCCCGCAATGTGCC..... AAAGT 2082  
392 lLutIleGluAsnValAlaAlaValAlaAspThrThrAsnGluLeuLysIle 408  
2083 GAACGCGATTGCATTTGACCAATACACCCCAAGCACT... TTGGTGT 2129  
409 GluLysProAspThrIleLysGlyGluAspAspAlaGluArgLeuIuLys 425  
2130 CGCACCCGATCAAAAGCCATCATCTGTACACTGTGACCTGACGACGAGTC 2179  
425 sGluProLysLysAlaValSerAspAspSerGluSerLysGluAlaSerP 442  
2180 TGACAAATTTGTCGAANAANAANCATTACGACAGCATTAAGTACTTCA 2229  
442 rGluGlyGlnAlaGluProGluProLysAspGluThrValAspValGlu 458  
2230 TTG..... ACTAAGACNGACNTAAGCGGACANTGNTAGNTNNCCA 2270  
459 MetLysMetAlaThrSerGluAspGluGluProMetThrGluLeuProAr 475  
2271 TNACGNTNTNTTAAANCTCNCNGGCGNTGNCNCACTNAGGCAATCTTA 2320  
475 gLleThrAsn..... AlaValAsnGlyAspLeu 484  
2321 GTGCAATGGCGATACGTTTATACAGTCAGCCACCAACGCCACCCAA... 2367  
485 ...AsnGlyAspLeuLysAlaSerIleGlyLysProLysSerLysPro 499  
2368 ...AACGGCAACCTAGC..... CTGTCGGCAATGC 2396  
500 LysProLysAlaLysLeuSerSerIleIleGlnLysLeuIleAspSerVa 516  
2397 CCAGACCAACATTTATCAAGCCACATTAACGCGACACNCATCGGNTTCG 2446  
516 lProAlaArgLeuGluGlnMetSer..... 524  
2447 GCAATGCTTCATTTATATTAAGCAACAACGCGCGCAACAAAGCGAGTCG 2496  
525 ..... LysThrSerAlaValIleAlaSerThrThr 534  
2497 AGCGTTTCGCAACGCTTAAGCAACGTAAGCCATTCGCTACGCAACG 2546  
535 ThrSerSerPargIleGlyGlyLeuSerHis..AlaLeuThrHis 550  
2547 CAATGCTCCCTAGCCGATTAAGCAGTA.....T 2575  
550 sLysValSerProProSerSerAlaThrAlaAlaGlyArgLeuValGluT 567  
2576 TCACATTTGAAAACAGCGGCTTTACCGGCAACACTACGCGACAGCAAGAN 2625  
567 yThrHisThrGln..... 570





617 hrserasnlysalalaspcluglyglyleuTyrCysglyaspasp 633  
1243 TTTCGGCTCTCG.....CCTGAACAACACGA 1268  
634 ValThrLeuThrAsnLeuThrGlyLysThrLeuPheGlnGluAsnSerSe 650  
1269 AACGTGGCAAGCGCGCGCTTCATATCAGTAGAAGACAGTACCTTACTT 1318  
650 rglulysnlsglyglyleuSerLeuAlaSerGlyLysSerLeuThrm 667  
1319 GG.....AAAGTAAAC 1329  
667 ethrSerLeuGluSerPheCysLeuAsnAlaAsnThrAlaLysGluAsn 683  
1330 GGC..... 1332  
684 GlyGlyGlyAlaAsnValProGluAsnLeValLeuThrPheThrTyrTh 700  
1332 ..... 1332  
700 rProThrProAsnGluProAlaProValGlnGlnProValTyrGlyGlu 717  
1333 .....GTGGCAACGACCGCCTGTCCAAATCGCAACGACGCTGCAC 1377  
717 lalauValThrGlyAsnThrAlaThrLysSerGlyGlyLysTyrThr 733  
1378 GTTCAGGCCAAGGGGAAACCAAGGCTCGATC..... 1410  
734 LysAsnAlaAlaPheSerAsnLeuSerValThrPheAspGlnAsnTh 750  
1411 .AGCGTGGCGCAGGTACAGTATTTGGATCAACGAGCAGACAGTAAAG 1459  
750 rSerSerLysnlsglyglylalaLeuLeuThrGlnLysAlaAlaAspLys. 766  
1460 GCAAAAAACAACCTTTAGTGAATCGGCTGTGNTCAGCGGAGGGGTACG 1509  
767 ..ThrAspCysSerPheThrTyrIle..... 774  
1510 GTGCACATGATGCCGATATCAGTATTCACCCGACAACTCTATTTCGG 1559  
775 ThrAsnValAsnIleThrAsn.....AsnThrAlaThrGln 786  
1560 CTTTCGGCGCGGACGTTGGATTTAAAGGGCATTCGCTTCGTTCCACC 1609  
786 YAsnGlyGly.....IleAlaGlyGlyLysAlaHisPheAsp 800  
1610 GTATTCAAAAT.....ACCGATGAAGGGCGCATG 1638  
800 rglleAspAsnLeuThrValGlnSerAsnGlnAlaLysLysGlyGly 816  
1639 ATTGCNATCATTAATGCCACAAACAATCCACGCTTACCATTTACAGGG.. 1686  
817 ValTyrLeuGluAspAlaLeuIleLeuGlnLysVal...IleThrGlySe 832  
1687 .....AATGAAGTATTACACAACGAGTGTAGATATCATTAAGACTTA 1732  
832 rValSerGlnAsnThrAlaThrGlnSerGlyGlyIle..... 845  
1733 ATTACAGCAAGAATTT.....GCCTACACGCTGGTTGGCGAG 1773  
846 ..TyrAlaLysAspIleGlnLeuGlnAlaLeuProGlySerPheThrIle 861  
1774 AAAGAT.....ACGACAAACGAAGCGCGGCT 1802  
862 ThrAspAsnLysValGluThrSerLeuThrThrSerThrAsn..... 875  
1803 CAACCTGTTTACAGCCCGCGCAGAGAAGACGACCCGCTGCTTTCCG 1852  
876 .....LeuTyrGlyG 879  
1853 GCGGACAATTTAAAGCAACATCAGCAAAACAACGCAACCTGTTT 1902  
879 lylglylIeTyrSerSerGlyAlaValThrLeuThrAsn..... 891

1903 TTCAGCGGACAGACCGACACCGCCTACATTCATTGGAAGCGGTG 1952  
892 lIeSerGly..... 894  
1953 GTCAAAAAATGGAAGTATCCCAAGAGAAATCGTGGACAAACGACT 2002  
895 .....ThrPheGlyIleThrGlyAsnSerValIle..... 904  
2003 GGATCAACCGCACGTTTAAAGCGGAAATTTCCATATTCAGGCGGG... 2049  
905 .....AsnThrAlaThrSerGlnAspAlaAspIleGlnGlyGly 918  
2050 ...CAGCGGTGATTTCCCGCAATGTGCCAAGTGAAGCGGATTCGCA 2096  
919 lIeTyrAlaThrThrSerLeuSerIleAsnGlnCysAsnThrProIleLe 935  
2097 TTTCAGCAATCAAGCCACGAGTTTGTGTGCGACCGCATCAAGCC 2146  
935 upheSerAsnAsnSerAlaAla..... 942  
2147 ATACAAATCTGTACAGTTGCGACTGACNGGTCTGACAAATTTGTGCA 2196  
943 .....ThrLysThrSerThrThrLys 950  
2197 NAANCATTAACGACGATTAAGTATGCTTCATTCATTAACGACGACMT 2246  
951 GlnIleAlaGlyGlyAlaIlePheSerAlaAlaValThrIleGluAsnAs 967  
2247 NAGCGGCATGTTNAGCNCTNCCATNACGNTNNTTAAANCCTCNGGCG 2296  
967 nserGlnProIleIlePheLeuAsnAsnSerAlaLysSerGluAlaThrT 984  
2297 NTGCNNCACTNAANGCAATCTTAGTCAAAATGGCATACAGTTATACA 2346  
984 hrAlaIaThrAlaGlyAsnLysAspSerCysGly..... 995  
2347 GTACGACACAGCGCACCCCAAAAGGCAACCTTACGCTGTGGCGAATGC 2396  
996 .....GlyAlaIleAlaAlaAsnSerValThrLeuThrAsnAsnTr 1009  
2397 CCAAGCAATTTAATCAAGCCACATTAAACGCAACGCAACGCGTTCGG 2446  
1009 ogIuIleThrPhe.....LysGlyAsnTyrAlaGluThrG 1021  
2447 GCAATGCT.....TCATTTAATCTAAGCAACACGCC..... 2478  
1021 lylGlyAlaIleGlyCysIleAspLeuThrAsnGlySerProProAlaGlys 1037  
2479 .....GCACAAACGCGAGTGTGACGCTTCGCAACGCTAAGGC 2519  
1038 ValSerIleAlaAspAsnGlySerValLeuPheGlnAspAsnSerAla 1054  
2520 AAC.....GTACGCAATTCG 2536  
1054 uAsnATGlyGlyAlaIleTyrGlyGluThrIleAspIleSerArgThg 1071  
2537 CACTCAAC.....GCCAT..... 2550  
1071 lylAlaThrPheIleGlyAsnSerSerLysAsnAspLysSerAlaIleCys 1087  
2551 .....GTTCGCTAGCGGATTAAGCGGATTCGCTTTGGAATA 2588  
1088 CysSerThrAlaLeuThrLeuAlaProAsnSerGlnLeuIlePheGln 1104  
2589 CAGCGCTTTACCGGACCAATCAGCGGCAAGAGAACGATTAACACT 2638  
1104 nAsnLysValThrGluThrThrAlaThrThrLysAlaSerIleAsnAsn 1121  
2639 TA.....AAAGACAGCAATGACGCTGCG 2664  
1121 euGlyAlaAlaIleTyrGlyAsnAsnGluThrSerAspIleThrIleSer 1137

```

2665 TCAGCAGCGAATTAGCAATTAAACCTTGACACGCCACATTACACT 2714
      ::::::::::: ||| ::::::::::: |||
1138 LeuSerAlaIleuAsnGlySerIlePhePheLysAsnAsnLeuCysThrAl 1154
      ::::::::::: ||| ::::::::::: |||
2715 CAATTCCCGCTATCGCCACGATGCTGCAGCGCGCAACCGCAGNGTGT 2764
      ::::::::::: ||| ::::::::::: |||
1154 aThrAsnLysIYrCysSerIleAla.....GlyAsnValL 1166
      ::::::::::: ||| ::::::::::: |||
2765 CAGACACGCCGCGCGCGCTTCGCGCTTATTCCTATTAATGACACCG 2814
      ::::::::::: ||| ::::::::::: |||
1166 ySpherThrAlaIleGluAlaSerAlaGlyLysAlaIleSerPheTyrAsp 1182
      ::::::::::: ||| ::::::::::: |||
2815 CCACCTTCGCTA.....CAATCCGCTTCAACAGCTGACGCTAA 2855
      ::::::::::: ||| ::::::::::: |||
1183 AlaValAsnValSerThrLysGluThrAsnAlaGlnGluLeuLysLeuAs 1199
      ::::::::::: ||| ::::::::::: |||
2856 CGGCAATTGTAACNGTCAAGACATTCGCTTATGTCGGAACCTTTC 2904
      ::::::::::: ||| ::::::::::: |||
1199 nGluLysAlaThrSerThrGlyThrIleLeuPheSerGlyLueuHISG 1216
      ::::::::::: ||| ::::::::::: |||
2905 .....GGCTACCGAAGCGCAAAATG..... 2925
      ::::::::::: ||| ::::::::::: |||
1216 LuAsnLysSerTyrIleProGlnLysValIThrPheAlaHisGlyAsnLeu 1232
      ::::::::::: ||| ::::::::::: |||
2926 .....AAGCTGCGCGAAAGTTCGAGAGNACTTACACC..... 2958
      ::::::::::: ||| ::::::::::: |||
1233 IleLeuGlyLysAsnAlaGluLeuSerValIserPheThrGlnSerPr 1249
      ::::::::::: ||| ::::::::::: |||
2958 ..... 2958
1249 oGlyThrThrIleThrMetGlyProGlySerValLeuSerAsnHisSerL 1266
      ::::::::::: ||| ::::::::::: |||
2959 .....TTGGCGGTCAACACATACCGCGCAACACCGCTAAGC 2994
      ::::::::::: ||| ::::::::::: |||
1266 ySGLuAlaGlyGlyIleAlaIleAsnAsn.....ValIle 1277
      ::::::::::: ||| ::::::::::: |||
2995 CTGATCAATTGACGCTAGTAGAGAGAAAGACAAC 3030
      ::::::::::: ||| ::::::::::: |||
1278 IleAspPheSerGluIleValProThrLysAspAsn 1289
      ::::::::::: ||| ::::::::::: |||
seq_name: SwissProt_40:OMP_RICJA
seq_documentation_block:
ID OMP_RICJA STANDARD; PRT; 1656 AA.
AC 006653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas5) (OMPb)
DE (Comp B) [contains: 120 kDa surface-exposed protein (Surface protein
antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptide].
GN OMPB
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-YH:
RA Uchiyama T.;
RT "Sequencing of the gene encoding the protein omp B of Rickettsia
japonica."
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
SIMILARITY).
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
(BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
LAYER WITH HEXAGONAL SYMMETRY.
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AB003681; BAA20138.1; -
DR InterPro: IPR003858; rompa_rompb.
DR Pfam: PF02708; rompa_rompb; 1.
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1338 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1339 1656 32 KDA BETA PEPTIDE.
FT DOMAIN 528 533 POLY-GLY.
FT SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;

```

## alignment\_scores:

```

Quality: 191.00 Length: 1512
Ratio: 0.284 Gaps: 73
Percent Similarity: 44.511 Percent Identity: 19.246

```

## alignment\_block:

```

US-09-303-518D-651 x OMPB_RICJA ..

```

```

Align seg 1/1 to: OMPB_RICJA from: 1 to: 1656

```

```

208 GATATTGAGGTGTACACAAAAAGGAGGAGTGTGCGCAATCAATGAC 257
      ||| ::::::::::: ||| ::::::::::: |||
347 AspGlyGlnValIleValIleValIleValIleValIleValIleVal 361
      ::::::::::: ||| ::::::::::: |||
258 AAAAGCCCGCATGATGATTTCT.....GTGGTGGCGCTAAGC 298
      ::::::::::: ||| ::::::::::: |||
361 nAlaAsnGlyGlnValAsnPheArgHisIleValAspValGlyIleAspG 378
      ::::::::::: ||| ::::::::::: |||
299 GCGTGGCGGATGTTGGCGATCAATATATGTTGACGTGCGACATPAC 348
      ::::::::::: ||| ::::::::::: |||
378 LyrThrThrAlaPheLysThrAlaAlaSerIleValAlaIleThrGlnAsn 394
      ::::::::::: ||| ::::::::::: |||
349 GCGGCTATATACAAAGTTGATTTGCT.....GCGAGAGGAAGNAA 389
      ::::::::::: ||| ::::::::::: |||
395 SerAsnPheGlyThrThrAspPheGlyAsnLeuAlaGlnValIleThyA 411
      ::::::::::: ||| ::::::::::: |||
390 TCCGATCAGCACCGCTTTTCTTACCAATGTGAAAGAAATATATATA 439
      ::::::::::: ||| ::::::::::: |||
411 IProAsp.....ThreThrIleuThrGlyAsnPheT 422
      ::::::::::: ||| ::::::::::: |||
440 AGCCTGACAAATTCACACCCCTTACACGCGCATTCATCATATGCGGCTTG 489
      ::::::::::: ||| ::::::::::: |||
422 hrgLysAspAlaAsnAsnProGlyAsn.....Thr 431
      ::::::::::: ||| ::::::::::: |||
490 CATATAATTGTTCACAGTGCAGAACTCTCGAATATGACGACATGAC 539
      ::::::::::: ||| ::::::::::: |||
432 AlaGlyValIleThrPheAlaIleAsnGlyThrLeuAlaSerValSerA 448
      ::::::::::: ||| ::::::::::: |||
540 GGGGAATACCTATTCGATTAAGAAATAATATCCGCGCGTCCGATCG 589
      ::::::::::: ||| ::::::::::: |||
448 IaSpAlaAsnValAlaValIleThrAsnAsnIleThrAlaIleGluAlaSer 464
      ::::::::::: ||| ::::::::::: |||
590 GCTCAGACACACCTATTTGGCGTTATGATGATG.....ACAACACGCG 633
      ::::::::::: ||| ::::::::::: |||
465 Gly.....ValGlyValAlaGlnLeuSerGlyThrIleThrAl 477
      ::::::::::: ||| ::::::::::: |||
634 GAT.....TTATCTACTCCGCGCA...TGCTTAATTTGGCGGCAATA 673
      ::::::::::: ||| ::::::::::: |||
477 acGluLeuArgGlyGlyAsnAlaGlySerValPheLysLeuAlaAspGlyT 494
      ::::::::::: ||| ::::::::::: |||
674 CACATATGACGGGTGGGAAATATGCGGTATTTGTTGACGCGCAT 723
      ::::::::::: ||| ::::::::::: |||
494 hValIleAsnGlyLysValAsnGlnThrVal.....LeuValGlyGly 508
      ::::::::::: ||| ::::::::::: |||

```



724 GTGCGCATGCCAAGCATTA.....GGCCCTATGCCGATTGGACG 764  
||| |||::: ||| |||::: |||  
509 ValLeuAlaIaIeIyAlaIeIhThrLeuAspGlySerAlaIhThIeThI 525  
||| |||::: ||| |||::: |||  
765 TGGCGGAGCGACAGCGGCTTCCCAATGTTATTATGACAAACAAACA 814  
||| |||::: ||| |||::: |||  
525 yAspIleGlyAsnGlyGlyGlyAlaIaIeIAsnIeThr... 540  
||| |||::: ||| |||::: |||  
815 ATTAATGCTGCTCAACGAGATTTCACAAACCGGCTACCTTATTCGGC 864  
||| |||::: ||| |||::: |||  
541 .....LeuAlaAsnAspAlaIhThrLysThr...LeuThrLeuGlyGly 553  
||| |||::: ||| |||::: |||  
865 AGGGAAGAGGTTTCCAGCTGATAGCAACAGATTGTTCTACATGACAT 914  
||| |||::: ||| |||::: |||  
554 Ala.....AsnI 556  
||| |||::: ||| |||::: |||  
915 TTACAGAGCGATACATACCTGCTNTTTGAACCGCGCAGTAAAGGAC 964  
||| |||::: ||| |||::: |||  
556 eIleSerAlaAsnGlyGlyThrIleAsnPhgI...AlaAsnGlyGlyT 572  
||| |||::: ||| |||::: |||  
965 ATTTTCCCTTTACATCCAAACAACAGGTACGGGTACGTAACAAACC 1014  
||| |||::: ||| |||::: |||  
572 hTrIleYsLeuThrSerThrGlnAsn.....AsnIleValIaIaSpCs 586  
||| |||::: ||| |||::: |||  
1015 AAGAAAGGTNNCCAAATCCAAAGCTTAAGTACAGACAGTCCACACTT 1064  
||| |||::: ||| |||::: |||  
587 AspLeuAlaIeIaIhThrAspIeIhThrGlyAlaIaIaSpAspSer 603  
||| |||::: ||| |||::: |||  
1065 TGACGAATCTTTGAATGAATGAATGAATGAACCGATTACCGCGCAGG 1114  
||| |||::: ||| |||::: |||  
603 uThrAsnAlaIaIhThrLeuThrIleSerGlyThrIleGlyIleIleGly 620  
||| |||::: ||| |||::: |||  
1115 GTTTATATAGTACCGTCCAAAGTTAAACAACGGTAAACCTTTCTTT 1164  
||| |||::: ||| |||::: |||  
620 LaAsnAsnThrThrLeuGlyGlnPhgAsnIleGlySerSerThrThr 636  
||| |||::: ||| |||::: |||  
1165 ATGATATACGCAAC.....GGCAACTCATCTTATCAACACAT 1205  
||| |||::: ||| |||::: |||  
637 LeuAsnGlyGlyAsnValaIaIeIAsnGlnIleuValIleGlyAsnAsnG 653  
||| |||::: ||| |||::: |||  
1206 CAAC...CAAGCGCGCGCGGCTTTGTTATTT..... 1233  
||| |||::: ||| |||::: |||  
653 ySerValGlnPhgAlaIaIaAsnThrThrLeuIleThrAlaIhThrAsn 670  
||| |||::: ||| |||::: |||  
1234 .....CAAGGTATTTTACGCTGCTGCT...GAACAACAAGAACG 1272  
||| |||::: ||| |||::: |||  
670 LaIaIaGlyGlnGlyIleIleIlePhgAsnProValIaIaAsnAsnThr 686  
||| |||::: ||| |||::: |||  
1273 TGGCAAGGCGGCGGCTTCATATCAGTAACAACGATCCGTTACTTGAA 1322  
||| |||::: ||| |||::: |||  
687 ThrLeuAlaIaIeIyThrAsnLeuGlySerAlaIaIaAsnProLeuAlaI 703  
||| |||::: ||| |||::: |||  
1323 AGTAAC.....GGCGTGGCAACAGCGGCTGTCGCAAAATCG 1360  
||| |||::: ||| |||::: |||  
703 uIleAsnPhgIleGlySerGlyAlaIaIaIaIaIaIaIaIaIaIaIaIaI 720  
||| |||::: ||| |||::: |||  
1361 GCAAGGC.....ACGCTGACGTTTCAAGCCAAAGGGAAGAAC 1398  
||| |||::: ||| |||::: |||  
720 yGlnGlyAlaI 736  
||| |||::: ||| |||::: |||  
1399 CAAGGTCGATCAGCGTGGCGGCGAGGTACATCTTTGGATCAGCAGCG 1448  
||| |||::: ||| |||::: |||  
737 ValGlySerPhg..... 740  
||| |||::: ||| |||::: |||  
1449 AGACGATAAGGCAAAAAACAAGCTTTAGTAAGATCGGCTGNTCAGCG 1498  
||| |||::: ||| |||::: |||  
741 .....ValPhgAsnAlaIeIyGlyLysAsnIleY 750  
||| |||::: ||| |||::: |||  
1499 GCAAGGAGTACGGTGCACACTGAATGCCGATTAATCAGTTCAACCCGCAAA 1548  
||| |||::: ||| |||::: |||  
750 aIleSerGlyThrValGlyGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnG 766  
||| |||::: ||| |||::: |||  
1549 CTCCTATTCGGCTTTCGGCGGCGACGTTTGGATTAAAGCGGCAATTCGCT 1598  
||| |||::: ||| |||::: |||

---

11 ||| |||::: ||| |||::: |||  
767 LeuAsp.....AsnGlyThrThrIa 773  
||| |||::: ||| |||::: |||  
1599 TTGCTTCACCGTATTCAAATACGATGAAGGCGGATGATGCTGCTC 1648  
||| |||::: ||| |||::: |||  
773 LysPhg.....LeuG 777  
||| |||::: ||| |||::: |||  
1649 ATAAAGCACAACACATCCACCGTTCACATTAACAGGAATGAAGTAT 1698  
||| |||::: ||| |||::: |||  
777 LysAlaIaIhThrPhgAsnGlyAsnThrThrIleAlaIaIaIaIaIaIaIaI 793  
||| |||::: ||| |||::: |||  
1699 ACACAACCGAGTGTAAAGATATCATATAGACTTAATTACGCAAGAAAT 1748  
||| |||::: ||| |||::: |||  
794 ...GlnIleSerGly.....AsnTyThrAlaIaIaIaIaIaIaIaIaIaIaI 803  
||| |||::: ||| |||::: |||  
1749 TGCTTACAC.....GGTTGGTTGGCGGCAAGATAGACACA 1786  
||| |||::: ||| |||::: |||  
803 eIleAlaIeIaI 820  
||| |||::: ||| |||::: |||  
1787 AAGCAACGGGCGGCTCAACCTTGTATTACAGCCGCCGACAGAACGCC 1836  
||| |||::: ||| |||::: |||  
820 rOIleAsnValThrLeuAsnGlyGlnAlaIaIaIaIaIaIaIaIaIaIaIaI 836  
||| |||::: ||| |||::: |||  
1837 ACCGCTGCTTCGCGC..... 1854  
||| |||::: ||| |||::: |||  
837 GlnIleThrValSerGlyProGlyAsnValIaIaIaIaIaIaIaIaIaIaI 853  
||| |||::: ||| |||::: |||  
1855 .GGAACAATTTAAAGCGCAACATCCAGCAACA..... 1887  
||| |||::: ||| |||::: |||  
853 nAlaGlyAsnTyThrIleGlyAlaIaIaIaIaIaIaIaIaIaIaIaIaIaI 870  
||| |||::: ||| |||::: |||  
1888 .....AAGGCAACGCTTTTTC.....AGCGGACAGCCGACACCGAC 1926  
||| |||::: ||| |||::: |||  
870 eIleSerLeuGlyAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaI 883  
||| |||::: ||| |||::: |||  
1927 GCTTACATCATTTAGAAC..... 1947  
||| |||::: ||| |||::: |||  
884 ...PhgAsnAspAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaI 899  
||| |||::: ||| |||::: |||  
1948 .GGGTGCTCAAAATGCAAGT..... 1968  
||| |||::: ||| |||::: |||  
899 lGlyAsnGlyThrAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaI 916  
||| |||::: ||| |||::: |||  
1969 .....ATCCACAGAGGAATGCTGTGGAC...AACGAC 2001  
||| |||::: ||| |||::: |||  
916 lYValaI 932  
||| |||::: ||| |||::: |||  
2002 TGATCAACCGCACGTTTAAAGCGGAAAT.....TT 2033  
||| |||::: ||| |||::: |||  
933 lIleValGlyLeuGlyLeuGlySerAspAsnGlyIleIleValaIaIaIaIaI 949  
||| |||::: ||| |||::: |||  
2034 CCATATTCAGGCGGCGGCTGATTTCCGCGCATGTGCGCAACGCG 2083  
||| |||::: ||| |||::: |||  
949 rThrLeuThrAlaI 966  
||| |||::: ||| |||::: |||  
2084 AAGCGCATTCGATTTGAGCAATCACGCCCAACGAGTTT..... 2124  
||| |||::: ||| |||::: |||  
966 hTrLeuSerGlyGlyAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaI 982  
||| |||::: ||| |||::: |||  
983 ThrGlyIleGlyAlaI 999  
||| |||::: ||| |||::: |||  
2172 GACGCTGTCGACAAATTTGTCGAAANAANCAATTACGACGATAAGGA 2221  
||| |||::: ||| |||::: |||  
999 rAsnAsnLeuGlyAsn.....IleI 1006  
||| |||::: ||| |||::: |||  
2222 TTGCTTCATTTGACTAAGACACNTNAGCGGAGTGAAGCTNNACNT 2271  
||| |||::: ||| |||::: |||  
1006 LeuAlaIhThrAsnThrThrIleAsn..... 1013  
||| |||::: ||| |||::: |||  
2272 MACGNTNNNTNAANCTCNCGGGCTGACNNCACTNAAGCAATCTTAG 2321  
||| |||::: ||| |||::: |||



```

4006 .....CAAAAGCGATTACCGCTACGAAACGTCA 4036
1481 aaanglyAsnMetSerLySglnIleAlaIaGlyThrMetAsnIleThrAspMet 1498
4037 ATATGCCACACCCCGCTTGGCTTACACGNTACCGNCGGCATTAA 4086
1498 hr.....PheGlyGlyAsnLeuThr 1504
4087 GCAGATTATTCATTCAAACCGCGCAACATTCATTCACACCTTATTT 4136
1505 ValGlyTyrAspTyrAsnIleMetGlnGlyValIleValThrProMetAl 1521
4137 MAGCCGTCTTATACCGATCCGCT..... 4161
1521 aglyLeuSerTyrLeuLysSerSerAspGlnAsnTyrLysGlnThrGlyT 1538
4162 .....TCGGCAAGTCCGCAACCGCTCAAT 4188
1538 hrThValAlaAsnLysGlnValAsnSerLysPheSerAspArgThrAsp 1554
4189 ACCGCGTATTCGCTCAGGATTTCGGCAAAACCGCCAGT...GGCGAATG 4235
1555 LeuIleValGlyAlaLysValAlaGlyThrMetAsnIleThrAsp 1571
4236 GGGCGTAAAGCCGCAATCAAGTTTCACGCTG 4269
1571 eAlaValTyrProGlnValHisAlaPheValVal 1582

```

seq\_name: SwissProt\_40:OMP\_RICPR

seq\_documentation\_block:

```

ID OMPB_RICPR STANDARD; PRT; 1643 AA.
AC 053020: 09ZCM0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scs5) (OMPb)
DE antigen (110 kDa outer membrane protein ompb); 32 kDa beta peptide).
GN OMPB OR SPAP OR SPA OR RP704.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_taxid=782;
RN 1[1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-BREINL;
RX MEDLINE-91045972; PubMed-2122457;
RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;
RT "Characterization of the gene encoding the protective paracrystalline-
RT surface-layer protein of Rickettsia prowazekii: presence of a
RT truncated identical homolog in Rickettsia typhi.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).
RN 1[2]
RP SEQUENCE FROM N.A.
RC STRAIN-BREINL;
RA Moron C.G., Yu X.J., Walker D.H.;
RT "Sequence analysis of ompb of Rickettsia prowazekii".
RL Submitted (JUN-1999) to the EMBL/Genbank/DBD databases.
RN 1[3]
RP SEQUENCE FROM N.A.
RC STRAIN-MADRID E;
RX MEDLINE-9039499; PubMed-9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Scharitz-Ponten T., Alsmark U.C.M., Podowski R.M., Neeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
RN 1[4]
RP PARTIAL SEQUENCE.
RC STRAIN-BREINL;

```

```

RX MEDLINE-92114896; PubMed-1370573;
RA Ching W.M., Carl M., Dasch G.A.;
RT "Mapping of monoclonal antibody binding sites on CNBR fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii.";
RL Mol. Immunol. 29:95-105(1992).
RN 1[5]
RP IDENTIFICATION OF CLEAVAGE SITE.
RX MEDLINE-92104668; PubMed-1729180;
RA Hackstadt T., Messer R., Cieplak W. Jr.;
RT "Evidence for proteolytic cleavage of the 120-kDa outer
RT membrane protein of rickettsiae: identification of an avirulent mutant
RT deficient in processing.";
RL Infect. Immun. 60:159-165(1992).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPb FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M37647; AAA26390.1; ALT_INT.
DR EMBL: AF161079; MAD42234.1; -.
DR EMBL: A0235273; CAU15140.1; -.
DR InterPro: IPR003858; OMPA_rOMPb.
DR Pfam: PF02708; rOMP_rOMPb; 1.
KW Antigen; S-layer; Cell wall; Complete proteome.
FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.
FT VARIANT 257 257 32 KDA BETA PEPTIDE.
FT VARIANT 1010 1010 Y -> A (IN STRAIN BREINL).
FT VARIANT 1450 1450 A -> D (IN STRAIN BREINL).
FT VARIANT 178 179 AA -> VC (IN STRAIN BREINL).
FT CONFLICT 191 201 TQGEAPFLGA -> INKSSYHVS (IN REF. 1).
FT CONFLICT 212 212 T -> I (IN REF. 1).
FT CONFLICT 313 313 Q -> L (IN REF. 1).
FT CONFLICT 1104 1104 D -> G (IN REF. 2).
FT CONFLICT 1123 1123 T -> S (IN REF. 2).
SQ SEQUENCE 1643 AA; 169854 MW; 735FD392E6346CC CRC64;

```

alignment\_scores: Quality: 190.50 Length: 1376  
Ratio: 0.303 Gaps: 66  
Percent similarity: 45.712 Percent identity: 18.750

alignment\_block:

US-09-303-518d-651 x OMPB\_RICPR

Align seg 1/1 to: OMPB\_RICPR from: 1 to: 1643

```

661 ATATGCCGCAATACA.....CATATGCGGCTGGGAAT.. 696
257 ValGlyGlyAsnThrIleAsnPhsAsnGlyIleAspGlyThrGlyLysIle 273
697 .....ATATGCCGTAATTA 709
273 uValleuValSerLysAsnGlyAlaAlaThrGluPhsAsnValThrGlyT 290
710 GTTATGAGCGCGCATGCGCCCATGCCACGACTATGCGCCGATTT 759
290 hrLeuGlyGlyAsnLeuLysGlyIleIleGluLeuAsnThrAlaAlaVal 306
760 GCAGTGGCGGACGAGCGGCTTCGCCAATGTTATTATTCAGACAAC 809

```

307     | | | | |     : : : : :     315  
307     AlaGlyLeuIleSerGlnGly..... 315  
810     AAACAAATATGGCTGCTCAACGAGATTTTAAACCGGCTACCTTATT 859  
316     .....AlaAlaAsnAlaValIle..... 321  
860     CCGGCAGGAAAAACGGTTTCCAGCTGATACGAAAGATTGTTTACGAT 909  
322     ..GlyThrAspAsnGlyAlaGly.....ArgAlaIleGlyPheIleVal 335  
910     GACATTTTACAGAGCGATACACATACCGTCTNTTTGAAACCGGCGAGTA 959  
336     SerValAspAsnGlyAsnAlaIleThrIleSerGlyGlnValTyrAlaIy 352  
960     CGGACATTTTCTTACATCCACAACAACGCT..... 993  
352     SAsnMetValIleGlnSerAlaAsnAlaGlyGlyGlnValThrPheGlnH 369  
994     .....ACGGTACGGTACAGAAACCAAGAAAGTTNCCAAAT 1032  
369     IstIleValAspValGlyLeuGlyGlyThrThrAsnPheGlyThrAlaAsp 385  
1033     CCAAGCTTAAAGTA...CAGACAGTCCGACCTGTTGACGAATCTTGA 1079  
386     SerIysValIleIleThrGluAsnSerAsnPheGlySerThrAsnPheG 402  
1080     TGAACGTATAAAGAACCACTTACGCG..... 1107  
402     YAsnLeuAspThrGlnIleValIleProAspThrIysIleLeuIysGlyA 419  
1108     .....CGAGGGGGTGTAAATCAGTACCGCTCCAAAGTTAAACAACGTTGA 1152  
419     snPheIleGlyAspValIys.....AsnAsnGlyAsn 429  
1153     AACCTTTCTTTATCGATTAC...GGCAGCGCAAACTCATCTTATCAAA 1199  
430     ThrAlaGlyValIleThrPheAsnAlaAsnGlyAlaLeuValSerAlaSe 446  
1200     C.....AACATCAAC.....CAAG 1213  
446     rThrAspProAsnIleAlaValThrAsnIleAsnAlaIleGluAlaGlu 463  
1214     CGCGGGCGGGTGTGTATTTGAAGGTATTTACGGTCTCGCTGAAAC 1263  
463     IyAlaGlyValVal..... 467  
1264     AACGAAACGTGCAAGCGGGCGGCTCATATCAGTGA..... 1302  
468     .....GluLeuSerGlyIleHisIleAlaGluLeuArgIleuG 480  
1303     .....GACAGTACCGTTACTTGAAG 1324  
480     YAsnGlyGlySerIlePheIysLeuAlaAspGlyThrVal.....I 494  
1325     TAAACGGCGTGGCAACGACCGCTGTCCAAATTCGCAAAAGCAGCCTG 1374  
494     IeAsnGlyProValAsn..... 499  
1375     CACGTTCAAGCCAAAGGGAACCAAGGCTCGATCAGCGGCG..... 1419  
500     .....GlnAsnAlaLeuMetAsnAsnAlaLeuAlaIleGlySerI 514  
1420     .....CAGGTACAGTCAATTTGATCAGACGAGCAGACGATTAAGCA 1462  
514     eGlnLeuAspGlySerAlaIleIleThrGlyAspIleGlyAsnGlyIy 531  
1463     AAAAACAAGCCTTAGTGAATCGGCTTGTACGCGGAGGCTTACGCTG 1512  
531     AlaAsnAlaIleLeuGlnHisIleThrIleuAlaAsnAspAlaSerIysI 547  
1513     ...CACTGAATCCGATTAATCAGTTCACCCCGAC.....AAACT 1550  
548     LeuAlaLeuAspGlyAlaAsnIleIleGlyAlaAsnValGlyAlaI 564  
1551     CTAATTTGGCTTTGCGCGGAGCTTTGATTTA.....AAC 1588  
564     eHisPheGlnAlaAsnGlyGlyThrIleLysIleThrAsnThrGlnAsn 591  
1589     GGCATTCGCTTGTCCACCGATTCATAAATACGATGAAGGGCGATG 1638  
581     snIleValValAsnPheAspLeuAspIleThrThrAspIysThrGlyVal 597  
1639     ATTGNCNATCATTAATGCCACACACATCCACGCTTACATTCACAGGAA 1688  
598     ValAspAlaSerSerIleuThrAsnAsnGlnThrIleuThrIleAsnGlySe 614  
1689     TGAAGTATTACACACCGAGCTGTGAAGATATCAATAGCTTATATAC. 1737  
614     rIleGlyThrValAlaAlaAsnThrLysThrLeuAlaGlnLeuAsnIleG 631  
1738     .....AGCAAGAAATTCCTACACAGGTTGTGTTGGCAGAAAGAT... 1779  
631     LysSerSerLysThrIleLeuAsnAlaGlyAspValAlaIleAsnGlnLeu 647  
1780     ACGACCAAAACGAAACGGCGGCTCAACCTGTTTACACAGCC..... 1821  
648     ValIleGluAsnAsnGlySerValGlnLeuAsnHisAsnThrIyLeuI 664  
1822     .....GCCGCAAGAACCCGACCCNCGTGTCCGCGGAA 1858  
664     eThrLysThrIleAsnAlaAlaAsnGlnGlyGlnIleIleValAlaAla 681  
1859     CAATTTTAAACGGCAACATCAACGCAAAACGCAAACTGTTTTCAGC 1908  
681     spProLeuAsnThrAsnThrIleuAlaAspGly..... 692  
1909     GGCAGACGACACCGACCGCTACATCATTTAGAACGGGTGCTCAA 1958  
693     .....ThrAsnLeuGlySerAlaGluAsnPr 701  
1959     AATGAGAGTATCCCAAGCAAGAAATCGTGGGCAACAGCTGATC 2008  
701     OleuSerThrIleHisPheAlaThrLysAlaAlaAsnAlaAspSerIle 718  
2009     ACCGCAGCTTAAAGCGGAAATTCATATTCAGAGGCGGCGAGCGGTG 2058  
718     eAsnValGlyLysGlyValAlaAsnLeuTyrAlaAsn..... 730  
2059     ATTTCCCGCAATGTTGCCAAAGTGCAGAGCGGATTCATTTGACGAACA 2108  
731     IleThrThrAsnAspAlaAsnVal...GlySerIleuHisPheArgSerG 746  
2109     CGCCCAAGAGTTTGTGTT.....GTGCAACCGCATCAAAACCATACAA 2152  
746     yGlyThrSerIleValSerGlyThrValGlyGlyGlnGlnGlyHisIys 762  
2153     TCTGTACAGTTGCGACGTGACNGGTGTGCAAAATTTGTGCAANAANC 2202  
763     .....LeuAsnAsnLeuIleLeuAspAsn 770  
2203     ATTACCGAGATAAGTATGCTTCATTGACTAGACAGCATNTAAGCG 2252  
771     GlyThrThrValLysPheLeuGlyAspThrThr..... 781  
2253     CANTGTNAGNCTNCCNATNAGCNTNTNTAANAANCNCGGCGNTGCNN 2302  
782     .....PheAsnGlyGlyThrL 787  
2303     CACTNNAAGCATCTTAGTCNAATGGCGATACAGCTTATACAGTACG 2352  
787     yIleGluGlyLysSerIleLeuGlnIleSerAsnSnyrThrThrAsp 803  
2353     CAC.....AACGCACCAAAACGCAACCTTAGCTGTGGGCAATGC 2396  
804     HisValGluSerAlaAspAsnThrGlyThrIleGluPheValAsnThrAs 820

```

2397 CCAAGCAACATTTAATCAAGCCACATTTAAAGCGCAAC..... 2433
      ||| :||| ||| ||| :||| ||| |||
820 pProlethr.....ValthrLeuasnLysGlnGlyAlaTyrPheG 834
2434 .....NCATCGGNTTCGGCAATGCTTCATTTAAAT 2463
      ||| ||| ||| ||| ||| |||
834 lYValleuLysGlnValIleIleSerGlyProGlyasnIleValPheasn 850
2464 .....CTAAGCAAC..... 2472
      ||| ||| ||| ||| ||| |||
851 GluIleGlyasnValGlyIleValIleHisGlyIleAlaIleasnSerIleSe 867
2473 .....AAGCGCGCAAAAACGCGAGTCTGACGCTTTCGCAACAGCTT 2514
      ||| ||| ||| ||| ||| |||
867 rPheGluasnAlaSerLeuGlyThrSerIleuPheLeuProSerGlyThrP 884
2515 .....AAGCAACGTAAGCCATTCGCGACTCAAC 2544
      ||| ||| ||| ||| ||| |||
884 rLeuasnPValleuThrIleLysSerThrValGlyasnGlyThrValasp 900
2545 GGCATGTCTCCCTACCGCATAGCGATATTCATTTGAAACACGCGC 2594
      ||| :||| :||| :||| :||| :|||
901 ...AsnPheasnAlaProIleValValValSerGlyIleaspSerMetIl 916
2595 CTTTACCGGACAACTCAGCGGAGCAAGANACAGCATTCACCTTAA... 2640
      ||| ||| ||| ||| ||| |||
916 easnasnGlyIleIleGlyAspPylsLysasnIleIleAlaLeuSerL 933
2641 .....AAAGACAGCAATGACGCTCCCTCAGCGACGGAATTAGCAAT 2685
      ||| ||| ||| ||| ||| |||
933 euGlySeraspasnSerIleThrValasnAlaasnThrLeuTyrSerGly 949
2686 TTAACCTT.....GACAACGCGACCATTTACACTCATTCGCGCCA 2726
      ||| ||| ||| ||| ||| |||
950 lLeaGlyThrThrLysasnasnGlnGlyValThrPheThrSerGlyGly 966
2727 TCGCCACGATGCTGCAGCGCGCAACCGGACNGTGCAGACAGCGCCG 2776
      ||| ||| ||| ||| ||| |||
966 tProasnasn.....ProGlyThrIleTyrGlyLeuGlyL 978
2777 GCGCGCTTCGCGCGTTCCTATTATCGGTTACACCGCCAACTCGGTA 2826
      ||| ||| ||| ||| ||| |||
978 euGluasnLysSerProLysLeuLysGlnValThrPheThrThrAspTyr 994
2827 GAATCCCGT.....TTCAACAGCTGACGCTAAC..... 2856
      ||| ||| ||| ||| ||| |||
995 AsnasnLeuGlySerIleIleAlaasnValThrIleasnAspTyrVa 1011
2856 ..... 2856
1011 lThrleuThrThrGlyIleAlaGlyThrAspPheaspAlaLysIleT 1028
2857 .....GGCAAAATTGAACNGTCAGAGAACATTCGCTTATGTCGGAATC 2901
      ||| ||| ||| ||| ||| |||
1028 hrLeuGlySerValasnGlyasnAlaasnValArgPheValaspSerThr 1044
2902 TCGCGCTACCGACGAGCAAAATTGAAGTGGCGGAAGTTCGGAAGNAC 2951
      ||| ||| ||| ||| ||| |||
1045 PheSeraspProArgSerMetIleValAlaThrGlnAlaasnLysGlyTh 1061
2952 TTACAC.....TTGCGGTCACATACCGGC.....AAGC 2993
      ||| ||| ||| ||| ||| |||
1061 rValThrTyrLeuGlyasnAlaLeuValSerasnIleGlySerLeuasnPT 1078
2984 AACCCGTAGCCTCGATCAATTGACGTAGTGAA.....GGGAAA 3024
      ||| ||| ||| ||| ||| |||
1078 hrProValAlaSerValArgPheThrGlyasnAspSerGlyAlaGlyLeu 1094
3025 GACAACAAACGCGTTCGGAACCTTAATTC.....AC 3059
      ||| ||| ||| ||| ||| |||
1095 GlnGlyasnIleTyrSerGlnasnIleaspPheGlyThrTyrasnLeuTh 1111
3060 CCTCGAAAACGACACGTGATCCGCGCGCTGGCGTTACC..... 3100
      ||| ||| ||| ||| ||| |||
1111 rIleLeuasnSerasnValIleLeuGlyGlyGlyThrThrAlaIleasn 1127
3101 .....AATCATTCGCC 3111
1128 GlyIleIleaspLeuLeuThrAsnasnLeuIlePheAlaasnGlyThrSe 1144
3112 AAAGACGGGAGCTTCGCGCTGATATTCGGTCAAAAGCAAGAGCTTC 3161
      ||| ||| ||| ||| ||| |||
1144 rThrTrpGlyaspasnThrSerIleSerThrThrLeuasnValSer..Se 1160
3162 CGACAAACTCGGCAAGCGACGAAGCCAAAACAGCGGAA..... 3201
      ||| ||| ||| ||| ||| |||
1160 rGlyasnIleGlyIleValIleAlaGluaspAlaGlnValasnAlaT 1177
3202 .....AAAGCAACGCGCAAGC... 3219
1177 hrThrThrGlyThrThrThrIleLysIleGlnaspasnAlaasnAlaasn 1193
3220 .....CTTGACGCGCTGATTCGCGCGCGCGCGCATGCCGC 3254
      ||| ||| ||| ||| ||| |||
1194 PheSerGlyThrGlnAlaTyrThrLeuIleGlnGly.....Al 1207
3255 GGAAGAAGACAGAAAGCCTTGCCGAACGCGCGCGCGCGAGCGGAAA 3304
      ||| ||| ||| ||| ||| |||
1207 aArgPheasnGlyThrLeuGlyAlaProasnPheAlaValThrGlySera 1224
3305 ATGTGCGCATTTATCAGCGCGAGAGAAAGAAAACGGGTGCAGCGGAT 3354
      ||| ||| ||| ||| ||| |||
1224 snIlePheVal.....LysTyrGluLeuIleArg...AspSerasn 1236
3355 AAAGACAGCGCNTTGGCGAAACAGCGCGAAGCGGAAACCGCGCGGTAC 3404
      ||| ||| ||| ||| ||| |||
1237 GlnaspTyrValLeuThrArgThrAsnaspVal...LeuasnValThr 1252
3405 CACCGCTTCGCGCGCGCGCGCGCGCGCGGAGATTCCCGCAACCGC 3454
      ||| ||| ||| ||| ||| |||
1252 rThrAlaValGlyasnSerAlaIleAlaasn..... 1262
3455 AGCCCAACCGCAACCTCAACCCCAACCGGACGCGACTGATNAGCCGT 3504
      ||| ||| ||| ||| ||| |||
1263 .....AlaProGlyValSerGlnasnIleSerArgLysLeuGlySerThr 1277
3505 TATCCCATATAGCGGTTGATGATTTTCGCCACGCTCAACAGCGTTT 3554
      ||| ||| ||| ||| ||| |||
1278 Asn.....ThrAlaAlaTyrasnSerMetIle 1286
3555 CGCGCTACAGGACGAATTTGAC.....CGCGTGTTCGCG 3589
      ||| ||| ||| ||| ||| |||
1286 uLeuAlaLysaspProSeraspValAlaThrPheValGlyAlaIleAlaT 1303
3590 AAGACCGCGCGCAACGCTTGACAGACGCGCATCCGCGNACCAAAACAC 3639
      ||| ||| ||| ||| ||| |||
1303 hrAspThrSerAlaAlaValThrThrValasnLeuasnAspThrGln... 1318
3640 TACCGTTCGCAAGATTTCCGCGCTACCGGCACAAACGCACTGGCGCA 3689
      ||| ||| ||| ||| ||| |||
1319 ...LysThrIleasp..... 1322
3690 AATCGGTATGCAGAAAAACCTCGGACGCGCGCGCTGCGCATCTGTTT 3739
      ||| ||| ||| ||| ||| |||
1323 .....LeuLeuSerasnArgLeuGlyThrLeuArgT 1333
3740 CGCACACCGGACGAAACANTTGCAGACGCGCATCGGCAACTCGGCA 3789
      ||| ||| ||| ||| ||| |||
1333 yTrLeuSerasnAlaGluThr..... 1339
3790 CGGCTTGCCACGCGCGCGTTTTCGGCAATACGCGATCGGAGGTTGA 3839
      ||| ||| ||| ||| ||| |||
1339 ..... 1339
3840 CATCGCATTCAGACGCGCGCGGTTTTCAGACGCGGCAANTCTNACAGC 3889

```

```

1340 .....SerAspValAlaGlySerAlaThrGlyAlaValaSerSerc 1353
13890 GCATCGAGGCAAAATCCCGCGCGTGCATTCAGGCATTCAGCA 3939
1353 LyspGluAlaGluVal.....SerTyrGlyAlaTrrPaia 1364
3940 CGA.....TACCGCGCGGTTTCGGCGG 3962
1365 LysProHeterYrAsnIleAlaGluGlnAspLysGlyIleAlaG 1381
3963 ATTC.....GGCATCGAAACCGTAC.... 3981
1381 YTYrLysAlaLysThrThrGlyValValGlyLeuAspThrLeuAla 1398
3982 .....ATCGCGCGCAACGCCCTTTCGTC 4005
1398 eAspAsnLeuMetIleGlyAlaAlaIleGlyIleThrLysThrAsp 1414
4006 CAATAACGCGATTACGC.....TACGAACGCTCAATACGCAC 4046
1415 LysHisIleAspTyrLysGlyAspLysThrAspIleAsn..... 1428
4047 CCGCGTCTTGCCTTCAACCGTAC 4071
1429 ...GlyLeuSerHeserLeuTyr 1435

```

seq\_name: SwissProt\_40:YM96\_YEAST

seq\_documentation\_block:

```

ID YM96_YEAST STANDARD; PRT: 1140 AA.
AC 004893;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 113.1 kDa protein in PRES-FET4 intergenic region.
GN YMR317W OR YM9924.09.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
ON NCBI_taxid=4932;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RA Church C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (NOV-1995) to the EMBL/Genbank/DDJ databases.
CC -!- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC
CC EMBL: Z54141; CAA90835.1; -.
DR SGD; S0004936; YMR317W.
KW Hypothetical protein; Repeat.
SQ SEQUENCE 1140 AA; 113070 MW; 0153BEC42FE5427 CRC64;

```

alignment\_scores:

```

Quality: 186.00 Length: 1278
Ratio: 0.295 Gaps: 46
Percent Similarity: 49.374 Percent Identity: 18.310

```

alignment\_block:

US-09-303-518D-651 x YM96\_YEAST ..

Align seg 1/1 to: YM96\_YEAST from: 1 to: 1140

575 AGCGTGTCCGATCGCTCAGACACCACTATTGGCTTATGATGATGAC 624

```

24 SerThrThrSerThrThrThrProThrThrSerThrThrSerThrTh 40
625 AAACGCGCGATTTATCTACTCCGCCATGTTAATGGCGGCATAC 674
40 rSerThrLysValThrThrSerProGluIle.....IleValSerSers 55
675 ACATATGACGGTGGGGAATTAATGGCGTANTTACTTTCAGCGCGCATG 724
55 eSerThrLeuValSerSerValAlaProGluPheThrSerSerSers 71
725 TCGCCCATGCCAGCTATGCGCCCTATGCCGATTCGAGTGGCGCAGC 774
72 LeuSerSerAspThrIleAlaSerIle.....LeuSerSerGluSerIle 86
775 GACAGCGGTCCCAATGTTATTATGACAAACAAACAAATGAGCT 824
86 uValSerIlePheSerSerLeuSerTyrThrSerSerPheSerSert 103
825 GCTCAACGAGTTTACAAACGCGCTACCTTATTCGCGCAGG..... 868
103 hSerValAsnAspValGluSerSerThrSerglyProSerAsnSerTyr 119
869 AAACGCGTTTCAGCTGATACGCAAAAGATGCTTCTACGATGACATTTAC 918
120 SerAlaLeuSerSerThrAsnAlaGlnLeuSerSerSerThrThrGluTh 136
919 AGAGCGGATACACATACCGCTTNTTTGAAACCGCGCTACGATTCGACATTT 968
136 rAspSerIleSer.....SerSerAlaIleGlnThrSers 148
969 TTCCTTTTACATTCACAAACAGGTACGGGTACGCTTACAGAACCAAC 1017
148 eProGlnThrSerSerSerSerAsnGlyGlySerSerSertL.....PheIle 164
1018 GAAAGGTTTCCAAATCCAAAGCTTAAAGTACAGACAGTCCGCTGTTGA 1067
164 uGlyLysSerSerValLeuGlnThrThrAlaSerSerSerp..... 178
1068 CGATCTTTGAATGAACACTGATTAAGAACCAAGCTTACGGCGCAGGGGTG 1117
179 .....ThrThrAlaVal 182
1118 TTAATCAGTACCGTCCAGCTTAACACGCTGAACACCTTT...CTTT 1164
183 ThrSerSerThrPheThrThrLeuThrAspValSerSerSerpLysI 199
1165 ATCGATTACGGCAACGCGCAACGTCATCTATGCAACCAACATCAACCAAG 1214
199 eSerSerSerGlySerAlaValThrSerValGlyThrThrSerAspPhe 216
1215 CCGCGCGCGTGTGTATTGTAAGGTGATTTTACGGTCTCGCTGAANA 1264
216 eTrpLysValAlaPheSerSerSerThrSerAspValSerSerLeuSer 232
1265 ACGAAACGTGGCAAGCGCGCGCTTCATATCAATGCAACAGTACCGCT 1314
233 SerThrSerSerProAlaSerSerThrIleSer...GluThrLeuProH 248
1315 ACTTGGAAGTAAAGCGCGTGGCAACGACCGCTGCTCCAAATGCGGCA 1364
248 eSerSer.....ThrIleLeuSerIleThrSerSerProValSerSerc 263
1365 AGCAGCGTGCACGTTCAAGCCAAAGGGGAAACCAAGGCTCGATCAGCG 1414
263 LuAla.....ProSerAlaThrSerSerSerValSerSerc 274
1415 TGGCGCAGCGTACATGATTTTGATCAGACAGGCAAGCATAAAGCAAA 1464
275 GluAlaSer.....SerSerThrSerSerValSerSerc 286
1465 AAACAGCGCTTTAGTGAATCGGCTGNTCAGCGCGCAGGCGTACGGTGA 1514

```





```

3195 GCGGGAAGACACCGCCGACGCGGCGTGTGATTTGGCGCGGCG 3244
      |||
805 sPaYsSerGlyAsnSerSerThrAla..... 813
3245 GCGATGCCGCGGCAAGAAAGCGTTGCCGACCGCGCGCGCGCA 3294
      |||
814 .....SerArgIleAlaThrSerLeuProAsnLysThrThrPheVa 827
3295 GCGGGAAGAAATGTCGCATTATGACGCGGAGAGAGAAAAAGCGG. 3343
      |||
827 lSerSerLeuSerSerThrSerAlaHisAlaArgAsnIlePheAsnSer 844
3344 .....TGCAGGCGGATTAAGACAGCGCGTTCGCCAAACAGC 3379
844 hrValIleValSerArgIleThrAlaIleGlyThrValThrSerCys 860
3380 GCGAAGCGCAAAACCGCGGCGTACACCGCTTCCCGCGCGCGCGNC 3429
860 ..... 860
3430 GCGCGCGCGGATTTGCCGCAACCGCGACCGCAACCGCATC..... 3472
      |||
861 .....CysSerAsnProThrProAsnTyrAsnIleThrLysT 873
3472 ..... 3472
873 hrValIleValSerArgIleThrAlaIleGlyThrValThrSerCys 889
3473 .....AACCCACCGCGAGCGCGACCTGATNACCGCTTATGCC 3510
890 SerGlyGlyCysThrLysAsnArgLysSerThr..... 900
3511 AATAGCGGTTCAGTGAATTTCCGCCACGCTCAACAGCGTTTCGCCGT 3560
      |||
901 .ThrLeuIleThrIleThrAspIleAspAlaSerThrVal..... 913
3561 ACAGGACGAATTGGACCGCGTGTTCGCGAAG...ACCGCGCGCAACGCG 3607
      |||
914 .....ThrThrCysProGluLysGluValThrSerThrThr 925
3608 TTTGGACAGCGCATCCGACGACCAACCACTACCGCTTCGCAAGATTTC 3657
      |||
926 SerGlyAspGluAlaGluHisThrThrSerThrLysIleSerAsnPro 942
3658 CGCGCGCTCGCGCAACGCGACCGCGCAAAATCGTATGCGAAGAAA 3707
      |||
942 urThrSerThrPheSerGluSerPheLysAspMetLysThrSerGlu 959
3708 CCTCGGCGCGCGCGCGCTGCGCATCTGTTTTCGACAGACCGGA..... 3751
      |||
959 hrLysLysAlaLysProGlySerGluThrValArgSerSerSerPhe 975
3752 ...CGGAACATCTTCGACGACGCGCATCGCAACTCGCGCGCGTTC 3798
      |||
976 ValGluLysThrSerProThrThrLysAlaSerProSerThrSerPro 992
3799 CACGCGCGCGTTTCGCGCATTCGCGCATCGCGAGTTCGACATCGCGAT 3848
      |||
992 rGluSerLysAlaAlaGlyAsnThrSerValAlaThrAsnAlaSerPro 1009
3849 CAGCAGCGCGCGCGGTTTACAGCGGCGANTCTTCAGACGCGATCGAG 3898
      |||
1009 erThrSerProSerSerGluSerGlnGlyThrGlySerThrSerValGlu 1025
3899 GCAAAATCCGCGCGCGCGTGTGCTGCTACGCGCATTCAGCGACGATCC 3948
      |||
1026 AlaLysSerLysSerThr.....LysAsnSerG1 1035
3949 GCGGTTTCGCGCGCATTCGCGCATCGACATCGCGCGACGCGCT. 3997
      |||
1035 uGlyValSerThrThrLysAlaLysAsnThrSerThrValAlaLysSer 1052
3997 ..... 3997

```

```

1052 erThrGluSerProIleGlyArgGlyGluThrThrLeuGluThrIleIle 1068
3998 ATTTCGTCAAAAGACGATTAACCGCTACGAAAGCGTAAATATGCGCAC 4047
      |||
1069 ValSerSerGlnLysSerSerLeuThrSerGlnLeuSerSerSerThrG1 1085
4048 CCCGCTTCGCGTTCACCGCATCCGCGCGGCGGCGGCGGCGGCGGCGG 4097
      |||
1085 uLysValAlaAsnArgSerThrThrLysProThrAla.....AlaIleH 1099
4098 ATTCAACCGCGCGCACACATTCATCA 4126
      |||
1099 lGlyThrSerSerSerAlaLysGlnSer 1108

seq_name: SwissProt_40:PM21_CHLPN
seq_documentation_block:
ID PM21_CHLPN STANDARD; PRT; 1609 AA.
AC Q9Z6U5; Q9RB58;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable outer membrane protein pmp21 precursor (Polymorphic membrane protein 21).
DE PMP21 OR CPN0963 OR CP0897.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shihai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -! SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC (POTENTIAL).
CC -! SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AE001676; AAD19099.1; -
CC DR EMBL: AE002248; AAF38684.1; -
CC DR EMBL: AP002548; BAA99171.1; -

```



DR PNCI-2DPAGE; Q9RB58; -  
 DR TIGR; CP0897; -  
 DR InterPro: IPR003357; OMP.  
 DR Pfam: PF02385; OMP; 1.  
 KW Outer membrane; Signal; Multigene family; Complete proteome.  
 FT SIGNAL 1 30 POTENTIAL.  
 FT CHAIN 31 1609 PROBABLE OUTER MEMBRANE PROTEIN PMP21.  
 FT CONFLICT 420 420 I -> M (IN REF. 3).  
 SO SEQUENCE 1609 AA; 170865 MM; 2604C3E9FC4024CB CMC64;

## alignment\_scores:

Quality: 185.50 Length: 1849  
 Ratio: 0.245 Gaps: 88  
 Percent similarity: 40.941 Percent identity: 18.280

## alignment\_block:

US-09-303-518D-651 x PM21\_CHLPN ..

Align seg 1/1 to: PM21\_CHLPN from: 1 to: 1609

```

16 AACGGACACCGAAGACACCGGAAAGCCGGAACCGCCGCGATCCG 65
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4  LysylThrValArgSerTyrArgSer..... 13
66 CTTCGCGCTCTTACTTACCATATGCTGCTGCGCATTTCTCC 115
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
14 .PheSerHisSerValIleValAlaIleLeuSerAlaGlyIleLapheg 30
116 AAGCTGGGCGGACACACTTAT.....TTGCGCATC..AACTAC 153
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
30 IuAlaHisSerLeuHisSerSerGluLeuAspLeuGlyValPheAsnLys 46
154 CAATACATCCGCACTTGCC.....GAAATAAGGCAAGTTTC 194
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
47 GlnPheGluGluHisSerAlaHisValGluIuAlaGlnThrSerValle 63
195 AGTGGGCGGCAAGATATTGAGTNTACAAACAAAGGAGGTGTG 244
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
63 ulysGlySerAspProValAsnProSerGlnLysGluSerGluLysVal 79
245 GCAATACATGACAAAGCCCGCATGATGATTTTCTGCTGCTCC 294
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
80 .....LeuTyrThrGlnValProLeuThrGlnLysSer..... 90
295 AACGGCGTGGCGCATTTGGTGGCGCATATATATTGAGCGTGGACA 344
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
91 SerGlyLysLeuAspLeuAlaAspPheLeuGluHisPheGlu 107
345 TAACGGCGCTATACAACTGATTTTGGTGGCA.....G 382
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107 nHisLeuPheGluGluThrValPheGlyIleAspGlnLysLeuVal 124
383 GAAGNANTCCGATCAGCACCGTTTCTTACCAATTTGAAAGAAAT 432
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
124 rPheAspLeuAspThrArgAsnPheSer.....GlnPro 135
433 AATTATAGCTGACATTCACACCTTACAGCGCATTANCATATGCC 482
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
136 ThrGlnLysProAspThrSer..... 142
483 GCGTTTGCATTAATTTGTACAGATGCAGAACTGTCGAATGACGAG 532
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
143 .....AsnAlaValSerGluLysIleSerSerA 152
533 ACATGAGGGAATACCTATTCCGATTAAGAAATATCCGAGCGGTG 582
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
152 spThrLysGluAsnArgLysAspLeuGluThrGluAspProSerLys... 167
583 CGCATCGCTCAGACACCACTATTGCGTTATGATGATGACAAACG 632
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
168 .....LysSerGlu 170

```

```

633 C.....GATTATCTACTCCGCGCATGTTATTG 664
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
170 LysGluValSerSerAspLeuProLysSer..... 181
665 GCGCATACACATATGACAGGTGGGAAATATAGCGTANTTACTTTG 714
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
182 .....ProGluThrAlaValAlaAlaIle 189
715 ACCGGCGATGTGGCGCATGCGCAAGACTATGCGCTATG...CCGATTG 761
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
190 SerGluAspLeuGluIleSerGluAsnIleSerAlaArgAspProLeuGlu 206
762 AGGTGGCGGACGAGACAGCGCTTCCCAATGTTATTATGACAAACA 811
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
206 nGlyLeuAla.....PhePheTyrLysAsnThrS 216
812 ACAATAATGCTG.....CTCAAGGAGTTTACAA 843
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
216 eSerGlnSerIleSerGluLysAspSerSerPheGlnGlyIleIlePhe 232
844 ACCGGCTACCTTATTCCGCGACGGAAC.....GTTTCCAGCTGAT 887
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
233 SerGly.....SerGlyAlaAsnSerGlyLeuGlyPheGluAsnLe 246
888 ACCG.....AAAGATTGTTCT 904
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
246 ulysAlaProLysSerGlyAlaAlaValTyrSerAspArgAspIleVal 263
905 ACGATGACATTTACAGAGCGATACA.....CATACGCTCTTTTGA 948
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
263 heGluAsnLeuValLysGlyLeuSerPheIleSerGluSerLeuGlu 279
949 CCGCGAGTACGAGACATTTTCTTACATCCACAAACAGGTACGG 998
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
280 AspGlySerAlaAlaGlyValAsnIleValThrHisCysGlyAspVa 296
999 TACGTAACAGAAACCAAGAAAGTNTCCATCCAAAGCTTAAAGTAC 1048
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
296 LThrLeuThrAsp.....CysAlaThrGlyLeuAspLeuG 308
1049 AGACACTCCGACTGTTTGACGAATCTTTGAATGAATGATGAAGACA 1098
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
308 IuAlaLeuArgLeuValLysAsp..... 315
1099 GTTTACGCGGAGGGGTGTTATACGTACCTCCAGGTTAAACACG 1148
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
316 .....PheSerArgGlyValAla...ValPheThrAlaArgAsnHisGluVa 330
1149 TGAAGC.....CTTCTTTTATCGATTCAGGCAAG 1180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
330 GlnAsnAsnLeuAlaGlyIleLeuSerValValGlyAsnLysGlyA 347
1181 GCAACTCATCTTATCAACATCAACCAAGCGCGGCGTTTGTAT 1230
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
347 IalIleValValGluLysAsnSerAlaGluLysSerAsnLysGlyAlaPhe 363
1231 TTTGAAGTATTTTACGCTCGCTCGCTGAAACCAAGAAACGTCG 1280
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
364 AlaCysGlySerPheValTyrSer.....AsnAsnGlu..... 374
1281 CCGGGCGTTTATCATGTAAGACAGTACCTGCTTGAAGTAAGAAC 1329
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
375 .....AsnThrAlaLeuThrLysGluAsnG 383
1330 .....GCGTGGCAACGACCGCTGTCCAAATTCGC...AAA 1365
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
383 IuAlaLeuSerGlyAlaIleSerSerAlaSerAspIleAspIleGln 399
1366 GGCAGCGCTCAGCTCAAGCCAAAGGGAACCAAGGCTCGATAGCGT 1415
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
400 GlyAsnCysSerAlaIleGluPheSerGluAsnGlnSerLeuIleAla 416
1416 GGGCAGCGTACAGTCATTTTGGATCAGCAGGACAGCATAAAGCAAA 1465

```



2663 CGTCAGGACGGAATTA..... 2679  
|||||  
961 roserglyalaglyValGlnllealaaspolyglyThrValCysLeuGlu 977  
2680 ..... GGCATTAAACCTTGACAAAGC 2702  
|||||  
978 AlaPheglyglyAspIleLeuPhegluglyAsnIleasnPheaspGly.. 993  
2703 CACATTACATCAATCCGCTAT..... CGCAGATCCCTCAGCGC 2746  
: : : : :  
994 ..... SerPheasnAlaIleHisLeuCysglyAsnspserlyIleV 1008  
2747 CGCAACCGCGCAGNGTGTCAAGC..... 2769  
: : : : :  
1008 alGluLeuSerAlaValGlnaspIlysnIleIlePheGlnaspAlaIle 1024  
2770 ACGCCGCGCCCGCTTGGCCGCTTCCCTA..... TTATCCGTTACACC 2813  
||| : : : : :  
1025 ThrTyrgluGlnAsnThrIleArgglyLeuProaspIlysnAspValSerPr 1041  
2814 GCCAACTTCGGTAGAATCCCTTTCACACCGCTACGCTAACCGCAAT 2863  
| : : : : :  
1041 OleuSerAlaProSerLeuIlePheasnSerLysProGlnaspAspSerA 1058  
2864 TG...AACNGTCAAGAACATTCGCTTATGTGGAACCTCTTCGGCTAC 2910  
: : : : :  
1058 IaglnHisHisgluglyThrIleArgPheSerArgglyVal..... 1071  
2911 CGAAGCGCAAAATTGAAGCTGGCGGAAAGTTCCGAAGGNACTTACACCTT 2960  
||| : : : : :  
1072 ...SerIlyleProGlnIleAlaIleGlnIuglyThrLeuAlaIle 1087  
2961 GCGGCTCAC..... AATACCGGCA 2980  
| : : : : :  
1087 uSerGlnasnIagluLeuTrpLeuAlaglyLeuLysGlnIuglyThrGlys 1104  
2981 AGCAACCCGTA..... AGCTCGATCAATTGACGATGTGTA 3018  
: : : : :  
1104 erSerIleValLeuSerAlaGlySerIleLeuArgIlePheaspSerGln 1120  
3019 GCGAAGACAAACAAACCGCTG...TCCGAAACCTTATTTACCCCTGCA 3065  
: : : : :  
1121 ValAspSerAlaProLeuProThrGlnasnLysGlnIuglyThrLeu.. 1136  
3066 AAAGCAACACTCGATCGCGCGCGCTGCTTACCACACTCATCCGCAAG 3115  
||| : : : : :  
1137 .....ValSerAlaGlyVal..... 1141  
3116 ACGGGAGTTCGCGCTGCATATCCGTCGAAGAACAAGAG..... 3156  
: : : : :  
1142 .....GlnIleasnMetSerSerProThrProasnLysAspIlyAlaVal 1156  
3157 .....CTTCCGACAACTCGCAAGGCAGAA.....GCCAA 3188  
||| : : : : :  
1157 AspThrProValLeuAlaAspIleIleSerIleThrValAspLeuSerSe 1173  
3189 AAAACAGCGCGGAAAAAGCAACGCGCAAGCCTT.....GACGCGCTGA 3222  
: : : : :  
1173 rPheValProGlnIuglnaspIlyThrLeuProLeuProProGlnIleIle 1190  
3223 TTGCGCGCGCGCGCATGCCCGCAAAAGACAGAAAGCGTTGCCGAACCG 3282  
||| : : : : :  
1190 leProlysgly.....ThrIlysnHisSerAlaIleIleasp 1202  
3283 GCCCGCGCGGAGCGCGGAAAAATTCGCGCATTTATGACGCGGAGGAAGA 3332  
: : : : :  
1203 leuTyrsIleIleAspProThrAsnValGlyTyr.....GluAs 1215  
3333 GAAAAAAGGGGTGCGGCGGATAAAGACGCGCNTTGGCGAAACAGCGCG 3382  
: : : : :  
1215 nHisAlaLeuLeuSerSerHisLysAspIleProLeuIleSerLeuYst 1232  
  
3383 AACGGAAACCGCGCGGNTACC...ACGCGCTTCCCGCGCGCGCGC 3429  
|||||  
1232 hralaGluGlyMetThrGlyThrProThrAlaAspAlaSerLeuSerasn 1248  
3430 GCCCGCGCGGATTTGCCGCAACCGCGCCCAACG..... 3465  
: : : : :  
1249 IleLysIleAspValSerLeuProSerIleThrProAlaThrTyrglyH 1265  
3465 ..... 3465  
1265 sThrGlyValTPrSerGlnSerLysMetGluaspGlyAlaGluValaG 1282  
3466 .....CAACT.....CAACCCCAACCGCGCGCGCACTG 3495  
|||||  
1282 LyrTrpGlnProThrGlyTyrlLysLeuasnProGlnLysGlnAlaLeu 1298  
3496 ATNAGCCGTTATGCCAATAGCGGTTGATGATTTCCGCGACGCTCA 3545  
: : : : :  
1299 ValLeuAsnAsnLeuTrpSerHisTyrlThrAspLeuArgAlaLeuLysG 1315  
3546 CAGCGTTTTCGCGCTACGAGCAATTTGACGCGCGCTTGGCGAAGACC 3595  
: : : : :  
1315 nGluIlePheAlaHisHisThrIleAlaGlnArgMetGluLeuAspPhe 1332  
3596 GCCGCAACGCGNGTTTGACAAAGCNGCATCCGGAACAC..... 3633  
||| : : : : :  
1332 erThrasn...ValTrpGlySerGlyLeuGlyValValaGluaspCysGln 1347  
3634 .....AATACCTACCGTTCC..... 3648  
1348 AsnIleGlyLupheaspIlyPheLysHisIleuThrGlyTyrlaLe 1364  
3649 .....CAAGATTC..... 3657  
1364 uGlyLeuAspThrGlnLeuValGluaspPheLeuIleGlyGlyCysPhe 1381  
3658 .....CGCGCTTACCGCGCAACAAACGAC 3681  
1381 erGlnPhePheglyLysThrGlnSerGlnSerTyrlLysAlaLysasnAsp 1397  
3682 CTGCGCCAAATCGGTATGCAGAAAACTCGGACGCGCGCGCTGCGCAT 3731  
: : : : :  
1398 ValLysSerTyrl.....MetGlyAlaIleTyrlaGlyIle 1409  
3732 CCGTTTTCGCAACACCGGACGGAACAACTTCGACGACGCG...ATGC 3778  
||||| : : : : :  
1409 eleuAlaGlyProTrpLeuIleLysGlyAlaPheValTyrglyAsnIleA 1426  
3779 GCAACTCGGACGCGCTTGGCCACGCGCGCGTTTTCGGGCAATACGCGATC 3828  
: : : : :  
1426 snAsnaspLeuThrThrAspTyrglyThr..... 1435  
3829 GGCAGGTTGCACATCGCATACGACGCGCGCGGTTTTCAGCGCGCAN 3878  
: : : : :  
1436 .....LeuGlyIleSerThrGlySerTrpIleGlyLysGlyPh 1448  
3879 TCTNTCAGACGCGATCGAAGCAAAATTCGCGCGCGCTGCATTAAG 3928  
: : : : :  
1448 eIleAla..... 1451  
3929 GCATTTCAGGACGATACCGCGCGGTTTTCGCGGATTCGCGATCGAACCG 3978  
||| : : : : :  
1451 LyrThrSerIleAspTyrlArg..... 1457  
3979 TACATGCGCGCAACGCGCTATTTCGTCAA..... 4008  
|||||  
1458 TyrlleValasnProArgArgPheIleSerAlaIleValSerThrVala 1474  
4009 .....AAGCGATTAC...CGTACGAAACGTCATATGCGCA 4045  
: : : : :  
1474 lProPheValGluAlaGluTyrlValArgIleAspLeuProGlnIleSerG 1491  
4046 CCCCCGCTTTCGTTCAACCGNTACCGNGCGGCGCATTAAGCGATTAAT 4095

```

1491 LUGLNGLY.....LysGluValArg 1497
4096 TCATTCACACGGCGGACACATNTTCATCANCCTTATTNAGCGTTC 4145
1498 TTTTGGGGLYSTHrArgPheGluValAlaIleProPheGlyPheAl 1514
4146 CTATACCGATGCCGCTGGGGGAAAGTCCGAACAGCGCTCAATACCGG 4195
1514 aLeuGluHisAlaIleTyrSerArgGlySerArgAlaGluValAsnSerValG 1531
4196 TATTGGCTCAGGATTCGCGAAACCCGACGCG..... 4230
1531 InLeuAlaIleTyrValPheAspValTyrArgGlyProValSerLeuIle 1547
4231 .....GAATGGCGCGTAAACCGCGCA 4250
1548 ThrLeuLysAspAlaIleTyrSerTyrPlySerTyrGlyValAspIlePr 1564
4251 AATCAAGGTTTCACGCTGTCNCNTCCACGCTGCCGCGCAAGACCGCG 4300
1564 OCysLysAlaTrp.....LysAlaIleArgLeuSerAsnAsnThrG 1577
4301 AACTGGAAGCGGACACAGACCGCGGCAATCAATTAGCTACCGCTGG 4347
1577 IuTrpAsnSerTyrLeuSerTyrLeuAlaIlePheAsnTyrGluTrp.1592

seq_name: SwissProt_40:120K_RICRI
seq_documentation_block:
ID 120K_RICRI STANDARD; PRT; 1300 AA.
AC P14914;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 120 kDa surface-exposed protein.
GN P120.
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OX Rickettsiaceae; Rickettsiinae; Rickettsia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R;
RX MEDLINE=90136087; PubMed=2515418;
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT "Cloning, expression and sequence analysis of the gene encoding the
RT 120 kD surface-exposed protein of Rickettsia rickettsii."
RL Mol. Microbiol. 3:1579-1586(1989).
CC -1- FUNCTION: MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS
CC RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -1- MISCELLANEOUS: ITS C-TERMINUS POTENTIALLY MAY BEAR THE EPITOPES
CC -1- CONFERRING ANTIGENICITY TO THE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X16353; CAA34402.1; -
DR PIR: S07575; S07575.
DR InterPro: IPR003858; romPA_romPB.
DR Pfam: PF02708; romPA_romPB; 1.
KM Antigen: Glycoprotein; Cell wall; S-layer.
FT CARBOHYD 66 66 N-LINKED (POTENTIAL).
FT CARBOHYD 86 86 N-LINKED (POTENTIAL).
FT CARBOHYD 86 86 N-LINKED (POTENTIAL).

```

```

FT CARBOHYD 103 103 N-LINKED (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (POTENTIAL).
FT CARBOHYD 436 436 N-LINKED (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (POTENTIAL).
FT CARBOHYD 547 547 N-LINKED (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (POTENTIAL).
FT CARBOHYD 655 655 N-LINKED (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (POTENTIAL).
FT CARBOHYD 799 799 N-LINKED (POTENTIAL).
FT CARBOHYD 800 800 N-LINKED (POTENTIAL).
FT CARBOHYD 826 826 N-LINKED (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (POTENTIAL).
FT CARBOHYD 861 861 N-LINKED (POTENTIAL).
FT CARBOHYD 879 879 N-LINKED (POTENTIAL).
FT CARBOHYD 920 920 N-LINKED (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (POTENTIAL).
FT CARBOHYD 1116 1116 N-LINKED (POTENTIAL).
FT CARBOHYD 1128 1128 N-LINKED (POTENTIAL).
FT CARBOHYD 1140 1140 N-LINKED (POTENTIAL).
FT CARBOHYD 1146 1146 N-LINKED (POTENTIAL).
FT CARBOHYD 1211 1211 N-LINKED (POTENTIAL).
SQ SEQUENCE 1300 AA; 132801 MW; E09B52C3F647243D CRC64;

```

alignment\_scores: Quality: 185.00 Length: 1334  
Ratio: 0.322 Gaps: 58  
Percent Similarity: 43.103 Percent Identity: 18.441

alignment\_block:

us-09-303-518d-651 x 120K\_RICRI ..

Align seg 1/1 to: 120K\_RICRI from: 1 to: 1300

```

1138 TTAACACAGGCGTGAACCTTTCTTTATCGATTACGCAAC..... 1179
37 ILeThrGluAspSerAsnThrAlaIlePheGlyAsnLeuAlaI 53
1180 .....GCGAACTCATCTTATCAACACATC..... 1206
53 agInIleLysValProAsnAlaIleThrLeuThrGlyAsnPheThrGlyA 70
1207 .....AACCAAGC.....GCGGCGGTTTGAATTGAAGGTGAT 1242
70 spAlaSerAsnProGlyAsnThrAlaIleGlyValIleThrPheAspAlaAsn 86
1243 TTTACGCTCTCGCTGAAACGAAACGAGGCGGCGCTTCA 1292
87 GLYThrLeuGluSerAlaSerAlaAsp.....AlaAsnValAl 99
1293 TATCACTGAGACAGTACCTTACTTGAAAGTAAACGCGTGGCAACG 1342
99 aValThrAsnAsnIleThrAlaIle...GluLaserGlyAlaGlyVal 115
1343 ACGCGCTGTCGCAAAATCGGCAAGACGACGCTGACGCTTCAACCAAGG 1392
115 aGlnLeuSer.....GLYThrHisAlaIleGluLeuValGlu 127
1393 GAAACACAGCTCGATC...AGCGTGGCGAGGATACGATTTTGG 1439
128 GlyAsnAlaGlySerIlePheLysLeuAlaAspGlyThrValIle..... 142
1440 TCAGCAGGCGAGATAAAGCAAAACAGCCCTTACTGTAATCGCT 1489
143 .....AsnGlyLysValAsnGlnThrAlaLeuValGly 154

```

```
1490 TGNTCAGCGGAGGCGTACGTCAGTCGAATGCCGAATATCATGCTTCAAC 1539
      ::::: |||||::: ||:::
154 lYAlaLeuAlaAlaGlyThrIleThrLeuAspGlySerAlaThrIleThr 170
1540 CCCGAC..... 1545
      |||
171 GlyAspIleGlyAsnAlaGlyAlaAlaAlaLeuGlnArgIleThrIle 187
1546 ..... AACCTCTTTTCG 1558
187 uAlaAsnAspAlaIaLysThrLeuThrLeuGlyAlaAlaAsnIleIleG 204
1559 GCTTTCGGCGCGGACGCTTGATTTA..... AACGGCATTCGCTTTCG 1602
      || ||||| |||||::: |||||:::
204 lYAlaGlyGlyThrIleAspLeuGlnAlaAsnGlyGlyThrIleLys 220
1603 TTCCACCGATTCAAAT..... 1620
      ::::: |||||
221 LeuThrSerThrGlnAsnAsnIleValAlaAspPheAspLeuAlaIleAl 237
1621 .ACCGATGAAGGCGCATGATGNCATCATATGTCACAAACATGCCA 1669
      |||||::: ::::: ||::: |||||:::
237 aThrAspGlnThrGlyValAlaAspAlaSerSerLeuThrAsnAlaGlnT 254
1670 CCGTTACCATTAACAGGAGTAAGTAATACAAACGAGTGTGAAGAT 1719
      ||::: |||||::: |||||::: |||||:::
254 hrLeuThrIleAsnGlyLysIleGlyThrIleGlyAlaAsnAsnLysThr 270
1720 ATCATTAACATTAATTAC..... AGCAAAATAATGGCTTCAACGGTTG 1763
      ||::: |||||::: |||||::: |||||:::
271 LeuGlyGlnPheAsnIleGlySerSerLysThrValLeuSerAsnGlyAs 287
1764 GTTTGGCGAGAAAGAT..... 1779
      ::::: |||||
287 nValAlaIleAsnGlnIleuValIleGlyAsnAspGlyAlaValGlnPhe 304
1780 ..... ACGACCAAAACGACGCGCGCTC 1803
      |||||::: |||||::: |||||:::
304 lAhAspThrTyrlLeuIleThrIleThrIleThrAsnAlaAlaGlyGlnGly 320
1804 AACCTGTGTACAGCGCGCGGAGAGACCGCACCGCGCTTTCGCG 1833
      ::::: |||||::: |||||::: |||||:::
321 LysIleIlePheAsnProValAlaAsnAsnGlyThrThr... LeuAlaAl 336
1854 CGAACAATAATTAAACGC..... 1872
      ::||| |||||::: |||||:::
336 aGlyThrAsnLeuGlySerAlaThrAsnProLeuAlaGlnIleAsnPheG 353
1872 ..... 1872
353 lYSerLysGlyValAsnValAspThrValLeuAsnValGlyGlnGlyVal 369
1873 ..... AACATCAGCAAAACAAC..... GGCAAACT 1898
      |||||::: |||||::: |||||:::
370 AsnLeuTyrlaThrAsnIleThrThrThrAsnValGlySerPhe 386
1899 GTTTTACGCGGACCGACGACGACGCGCTACATCATTTAGAGCG 1948
      |||||::: |||||::: |||||:::
386 eValPheAsnAlaGlyGlyThrAsnIleValSerGlyThrValGlyGly. 402
1949 GGTGTCAAAAATGAGATATCCACAGAGAAATCGTGTGGACAC 1998
      ::::: |||||::: |||||::: |||||:::
403 ..... GlnGlnGlyAsnLysPheAsnThrValAlaIleLeuGlnAsn 415
1999 GACTGATTCNACGC..... ACGTTTAA..... 2022
      ::||| |||||::: |||||:::
416 GlyThrThrValLysPheLeuGlyAsnAlaIleThrPheAsnGlyAsnThr 432
2023 .... GCGGAAATTTCCATATTCAGGGCGGACGCGGTATTTCCGCA 2068
      ||| ||| ::||| ||| :::
432 rIleAlaAlaAsnSerThrLeuGlnIleGlyGlyAsnTyrlThrAlaAspC 449
```

```
2069 ATGTGCCAAAGTGAAGCGAT..... 2091
      |||||::: |||||::: |||||:::
449 ySValAlaSerAlaAspGlyThrGlyIleValGlnIlePheValAsnThrGly 465
2091 ..... 2091
466 ProIleThrValThrLeuAsnLysGlnAlaAlaProValAsnAlaLeu 482
2092 ..... TGNC 2095
482 sGlnIleThrValSerGlyProGlyAsnValAlaIleAsnGlnIleGlyA 499
2096 ATTGAGCAATCAAGCCAGAGTTTGTGTGGCGACCGCATTCAAAGC 2145
      ::::: ||||| |||||:::
499 snAlaGlyAsnHisHisGlyAlaValThrAspThrIleAlaPheGlnAsn 515
2146 CATACAATC..... 2154
      ::::: |||||
516 SerSerLeuGlyAlaValAlaPheLeuProArgGlyIleProPheAsnAs 532
2155 ..... TGTCACGCTTGGACGTGGACNGCTGTACAA 2185
      |||||::: |||||::: |||||:::
532 pAlaGlyAsnThrMetProLeuThrIleLysSerThrValGlyAsnLysT 549
2186 ATTGT..... GTCGAANA 2199
      ::||| |||||::: |||||:::
549 hrAlaLysGlyPheAspValProSerValAlaValLeuGlyValaAspSer 565
2200 ANCATTAACGAGATAAAGTATGCTTCATTGACTAAGACNGACNTNAG 2249
      |||||::: |||||::: |||||:::
566 ValIleAlaAspGlyGlnValIleGly..... AspGlnAs 577
2250 CGGCATGTAGACNTNCCATNACGNTNNTTAAANCTC..... 2289
      ::||| |||||::: |||||:::
577 nAsnIleValGlyLeuGlyLeuGlySerAspAsnGlyIleIleValaAsnA 594
2290 ..... NCGGGCMT 2298
      |||||::: |||||::: |||||:::
594 lAthrThrLeuTyrlaGlyIleSerThrLeuAsnAsnAsnGlnGlyThr 610
2299 GCNNCACTNNAANGCAATCTTACTGCAATGGGATACAGTAT..... 2343
      ::||| |||||::: |||||:::
611 ValThrLeuSerGlyGlyValProAsnThrProGlyThrValTyrlcyle 627
2343 ..... 2343
627 uGlyThrGlyIleGlyAlaSerLysPheLysGlnValThrPheThrThra 644
2344 ..... ACAGTCAGCCACACGACCGCACCCAAACGCGC 2373
      ::||| |||||::: |||||:::
644 sPTyrAsnAsnLeuGlyAsnIleIleAlaIleThrAsnAlaThrIleAsnAsp 660
2374 AACCTTAGCCTGTG..... GGCAATGCCCAAGCAACATTTAATCAACG 2417
      ::||| |||||::: |||||:::
661 GlyValThrValThrThrGlyGlyIleAlaGlyIleGlyPheAspGly 677
2418 CACATTAACCGCAACNCAFCGNTTGGGCAATGGTCTATTAT..... 2463
      |||||::: |||||::: |||||:::
677 sIleThrLeuGlySerValAsnGlyAsnGlyAsnValaIlePheAlaAspG 694
2464 .... CTAGCAACAACGCGGACCAAAACGCGAGTGTGAGCTTTCGAC 2508
      |||||::: |||||::: |||||:::
694 lYIleuSerAsnSerThrSerMetIleGlyThr... ThrLysAlaAsn 709
2509 AACGCTAAGCAACGTAAGCATTCGACATCAACGCAATGTCTCCT 2558
      |||||::: |||||::: |||||:::
710 AsnGlyThrValThrTyrlLeuGlyAsnAlaPheValGlyAsnIleGlyAs 726
2559 AGCGGATTAAGCAGATTCATTGAAACAGCGGCTTACGCGCAAC 2608
      ::||| ||| ::||| ||| :::
726 pSerAspThrProVal..... AlaSerValArgPheThrGlySerA 740
2609 TCAGCGGACAGCAAGANACAGA..... 2631
```

|||||: : :  
740 spserglYAlaIglYleuInglYAsnIleYrSerGIInValIleAspPhe 756  
|||||: : :  
2632 .....TTACCTTAAGACAGCGAATGACGCTGCCGTCAGG 2669  
||| : : : ||| : : :  
757 gLYthrYAsnleuInglYleValAsnSerAsnIleIleuInglYglY 773  
|||||: : : : :  
2670 CAGCGAATTGAGCAATTAAACCTTGACACGCGACCATTAACAATTA 2719  
||| : : : : :  
773 YthThAlAlIleInglYlYsIleAspLeuValThrAsnThIleuThrP 790  
|||||: : : : :  
2720 CCGCCTATCGCAGATGTCAGGCGCGCAACCGCGCAGNtGTCAAGAC 2769  
||| : : : : :  
790 he.....AlaserGIYthrSerThrTPglYAsnAsn... 800  
||| : : : : :  
2770 ACGCCGCGCGCGTTCGCGCGTCCCTTATTCGCTACACGCCAAC 2819  
||| : : : : :  
801 .....Th 801  
2820 TTGGTAGAATCCGCTTCAACAGCTGACGTA...ACGCGAATTTGA 2866  
|||||: : : : :  
801 rSerIleGIthr.....ThIleuThrleValAsnGIYAsnIleGI 815  
||| : : : : :  
2867 AC.....NGTCAAGAACATTCGCTTATGTGGAATCTCTC 2904  
||| : : : : :  
815 lYhISleValIleleuInglYAla..... 823  
2905 GCGTACCGAAGCGCAAAATTGACCTGCGGAAAGTTCCAGAGNACTTA 2954  
||| : : : : :  
824 .....GlnValAsnThrThrThrThrThrThrThrThrThrThr 833  
2995 CACCTTGGCGGTCAACATACCGGCAAGAACCCGTAAGCCTGATCAAA 3003  
|||||: : : : :  
833 rThrIleYsValGIAspAsnAlaAsnAlaAsnPheserGIYthrGIInT 850  
||| : : : : :  
3004 ..TTGACGTAAGTGAAGG.....AAGACAAACAAACCGCTG...TCC 3042  
||| : : : : :  
850 hrTYrThrleuIleInglYglYAlaIarGPhesInglYThrleuInglYser 866  
3043 GAAACCTTAATTACCCCTGCAAAAGACGACCTGATGCGGCGGTG 3092  
||| : : : : :  
867 ProAsnPhleAlaValThrGIYSerAsnIarGPhleValAsn..... 879  
3093 GCGTACCACTCATTCGCAAGACGCGAG..... 3123  
||| : : : : :  
880 ....TySerIleuIleIarGAlaIaAsnGIAspTYrValIleThrIarGT 895  
3124 ..TTCCGCTGTCAATATCCGCTCAAGAACAGCTTTCGACAAATC 3171  
||| : : : : :  
895 hrAsnAsnAlaGIAsnValValThrAsnAspIleAlaAsnSerProPhe 911  
3172 GCGAAGGAGAACCCAAAAACAGCGGAAAAAACAAACAGCGCAACGCT 3221  
||| : : : : :  
912 gLYglYAlaProglYValAsp.....GlnAsnIva 921  
3222 TGACGCGTGAATTGCGCGCGCGGATGCCCGGAAAAAGACAAAGCG 3271  
||| : : : : :  
921 lThrThrPhleValAsnAlaIthrAsnThrAlaIaIeTYrAsnAsnleuI 938  
3272 TTGCGGAA...CCGCGCGCGCGCGAGCGGGGAAATGTGCGCATTAATG 3318  
||| : : : : :  
938 euAlaIYAsnSerAlaAsnSerAlaAsnPhleValGIYAlaIleValThr 954  
3319 CAGCGCGAGAGAGAAAAACGGGTGACGCGGATTAAGACAGCGCNTT 3368  
||| : : : : :  
955 AspThrSerAlaAlaIleThrAsnValGIInleuAsp.....Le 967  
3369 GCGCAAGACAGCGGAGAGCGGAAACCGCGGNTACACCGCTTCCCCC 3418  
|||||: : : : :  
967 uAlaIYAspIleGIInAlaGIIn.....leuGIYA 977  
3419 GCGCGCGCGCGCGCGGATTTGCCGCAACCGCGAGCGCCCAACCGCA 3468  
||| ||| ||| : : : : :  
977 snarGIeGIYAlaIeuarGTyrIleuInglYthrProGIuThrAlaGIleuT 993  
3469 CCTCAACCCCAACCGCACCGGACCTGATMNCGCTTATTCGAATAGCGG 3518  
|||||: : : : :  
994 AlAGIYProGIu..... 997  
3519 TTTGAGTGAATTTTCCGCGACGCTCAACAGCGTTTTCGCGTACAGAGG 3568  
||| : : : : :  
998 .....AlaGIYAlaIleSerAlaIaIaValAlaIaIaGIYAspG 1010  
3569 AA...TTGACGCGCGCTGTTGCCGAAGACCGCGCAACGCGNtGTGACA 3615  
||| : : : : :  
1010 lAlaIleAspAsnVal.....AlaTYrGIYleThrPalA 1021  
3616 AGCNGCATCCGCGACCAACCAACACTAC...CGTTCGAAGATTTCCGCGC 3662  
||| : : : : :  
1022 lYsProPhleTYrThrAspAlaIleGIInSerIYsIYsGIYglYleuAlaGI 1038  
3663 CTACGCGCAACAAACCGACCTGCGCAATTCGGTATGCAAGAAAACTCG 3712  
||| : : : : :  
1038 YTYrIYsAlaIYsThrThrGIYValIleGIYleu..... 1050  
3713 GCAAGCGCGCGCTCGGACATCTGTTTTCGCAACCGGACGCAAAACANC 3762  
1050 ..... 1050  
3763 TTGACAGCGCATCGCAACTGCGACGCGCTTTCGCCAGCGCGCTTTT 3812  
||| : : : : :  
1051 ....AspThrleuAlaAsnAspAsnleuMetIleGIYAlaIaIle... 1064  
3813 CCGGCAATGAGCGCATGCGGCTGACATCGGATCAGACAGCGGCGCGG 3862  
||| : : : : :  
1065 .....GIYleThrIYsThrAspIle.....lYsInGIInA 1075  
3863 GTTTTACAGCGCGCANTCTNCAAGCGGATGCGGCAAAATCCGCGCGC 3912  
||| : : : : :  
1075 sPTYrIYsIYsGIYAspIYsThrAspValAsnGIYPheserPheserIleu 1091  
3913 CCGGTCTCATTTACGCAATTCAGCGACGA.....TACCGCGC 3950  
||| : : : : :  
1092 .....TyrlYAlaGIInGIInleuValIYAsnPhleAlaGI 1104  
3951 CCGTTTCCGCGATTCGCGATGCAACCGTACATCGC...GCAACGCGT 3997  
||| : : : : :  
1104 nGIYSerAlaIlePheserIleuAsnGIInValIYsInIYsSerGIInIarGT 1121  
3998 ATTTGCTC.....CAAAAACGGATTAACGCGC 4023  
||| : : : : :  
1121 YrPhleAspAlaAsnGIYAsnMetSerIYsGIInIleAlaIaIaGIYhIS 1137  
4024 TACGAACGTCATATTCGCCACCCCGGCTTTCGCTTCAACCGNTACCG 4073  
||| : : : : :  
1138 TyAspAsnMetThr.....PheGI 1144  
4074 NCGGCGCATTAAGCGCAGATTATTCATTCAACCGGCGCAACATNTTCA 4123  
||| : : : : :  
1144 YGIYAsnleuThrValGIYTYrAspTYrAsnAlaMetGIInGIYValleuV 1161  
4124 TCACNCTTATTNAGCCTGTCTTATTCAGATGCGCTTCGGGCAAGTIC 4173  
||| : : : : :  
1161 alThrProMetAlaGIYleuSerTYrleuIYsSerIarSpGIInIYr 1177  
4174 CGAACACGCGTCAATACCGCGNTATTGGTCAGATTTGGCAAA...AC 4220  
||| : : : : :  
1178 lYsGIuThrGIYThrThrValAlaAsnIYsGIInValIaIaIaSerIYsPhe 1194  
4221 CCGCAGTGCAGATYGGGCGTAAACCGCAATCAAGTTTCACGCTGT 4270  
||| : : : : :  
1194 rAspArGTThrAspIleIleValGIYAlaIYsValAlaGIYSerThrMeta 1211  
4271 CC 4272  
1211 sn 1211

```

seq_name: swissprot_40:WAPA_BACSU
seq_documentation_block:
ID WAPA_BACSU STANDARD: PRT: 2334 AA.
AC Q07833.1
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Wall-associated protein precursor.
GN WAPA OR N17G.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=93302506; PubMed=8316082;
RA Foster S.J.;
RT "Molecular analysis of three major wall-associated proteins of
RT Bacillus subtilis 168: evidence for processing of the product of a
RT gene encoding a 258 kDa precursor two-domain ligand-binding
RT protein."
RL Mol. Microbiol. 8:299-310(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=95219088; PubMed=7704263;
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT genome containing the hut and wapa loci."
RL Microbiology 141:337-343(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=97124196; PubMed=8696509;
RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
RA Miwa Y., Fujita Y.;
RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
RT containing the hic and cel loci, and creation of a 177 kb contig
RT covering the gut-sacx region."
RL Microbiology 142:3113-3123(1996).
CC -1- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
CC MOTILITY, SECRETION OR DIFFERENTIATION.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
CC INTO THE MEDIUM.
CC -1- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
CC MOTIF REPEATED 31 TIMES.
CC -1- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
CC SIMILARITY TO THE REPEAT IN E. COLI RBS GROUP OF PROTEINS (RBSA-D).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L05634; AAA22883.1; -
DR EMBL: D31856; BAA06656.1; -
DR EMBL: D29985; BAA06260.1; -
DR EMBL: D83026; BAA11683.1; -
DR EMBL: Z89124; CAB15959.1; -
DR PIR: S32920; S32920.
DR Subtilisin; BG10797; wapa.
DR InterPro: IPR003305; CBD_6.
DR Pfam: PR02018; CBD_6.1.
KW Cell wall; Repeat; Signal; Complete proteome.
FT SIGNAL 1 28 OR 32 (POTENTIAL).

```

```

FT CHAIN 29 2334 WALL-ASSOCIATED PROTEIN.
FT DOMAIN 504 869 3 X 101 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 504 605 1-1.
FT REPEAT 606 736 1-2.
FT REPEAT 737 869 1-3.
FT DOMAIN 1021 2139 31 X 21 AA APPROXIMATE TANDEM REPEATS OF
FT REPEAT 1021 1040 X(4)-G-X(4)-(YF)-X-D-X(2)-G-X(4).
FT REPEAT 1041 1061 2-1.
FT REPEAT 1062 1082 2-2.
FT REPEAT 1083 1102 2-3.
FT REPEAT 1103 1128 2-4.
FT REPEAT 1129 1148 2-5.
FT REPEAT 1149 1169 2-6.
FT REPEAT 1170 1190 2-7.
FT REPEAT 1191 1218 2-8.
FT REPEAT 1219 1238 2-9.
FT REPEAT 1239 1265 2-10.
FT REPEAT 1266 1285 2-11.
FT REPEAT 1286 1305 2-12.
FT REPEAT 1306 1325 2-13.
FT REPEAT 1326 1345 2-14.
FT REPEAT 1346 1365 2-15.
FT REPEAT 1366 1385 2-16.
FT REPEAT 1386 1405 2-17.
FT REPEAT 1406 1425 2-18.
FT REPEAT 1426 1445 2-19.
FT REPEAT 1446 1465 2-20.
FT REPEAT 1466 1485 2-21.
FT REPEAT 1486 1505 2-22.
FT REPEAT 1506 1525 2-23.
FT REPEAT 1526 1545 2-24 (APPROXIMATE).
FT REPEAT 1546 1565 2-25.
FT REPEAT 1566 1585 2-26.
FT REPEAT 1586 1605 2-27.
FT REPEAT 1606 1625 2-28.
FT REPEAT 1626 1645 2-29.
FT REPEAT 1646 1665 2-30.
FT REPEAT 1666 1685 2-31.
SO SEQUENCE 2334 AA; 258329 MW; B75138CCD278BA3 CRC64;

```

```

alignment_scores:
Quality: 184.00 Length: 1242
Ratio: 0.333 Gaps: 65
Percent Similarity: 44.444 Percent Identity: 18.680

```

alignment\_block:

US-09-303-518d-651 x WAPA\_BACSU

Align seg 1/1 to: WAPA\_BACSU from: 1 to: 2334

```

160 TATCGGCACTTGGCGAATAAAGCAAGTTGCGAGTGGCGGCGAAGA 209
||| :||||| :|
988 TyleuGlndeuThrluThraAlaAspGlnPheIleuLysThrLysAs 1004
210 T..... 210
1004 pGlnThraAlaLysPheAsnLysLysLysLysLysLysLysValy 1021
211 .....ATTGAGCTTACACAAAGAGGAG 237
1021 aAlaSpGlnHisAsnAlaThrValThrThrLysAsnPlsAsnGln 1037
238 TTGGTGGCAATGCAATGCAAAAGCCGATGATTTTCGTGCT 287
||||| :| :|
1038 LeuThraAlaIleThraAlaSerGlyAlaGlyLysLeuThrPheThrLysAs 1054
288 GTCCGCTAACGGCGTGGCGGCAATGCTGGCGC..... 318
||||| :| :|
1054 pGlnAsnGlnHisValThrSerLethrLysProlLysAsnLysLysValT 1071
319 .....GATCAATATATTGAGCGGTGGCAAT...AACGGC 351
||||| :| :|

```

1071 hTyrSerTyrGluAsnSpleuLeuLysValThrAspThrAspGly 1087  
352 GCGTATACACAGTTGATTTGGTCGGAAGAAAGNAATCCGATACACA 401  
1088 ThrValThrSerTyrAspTyrAspSerGly 1098  
402 CCGTTTCTTACCAATGTGAAAAAATATATATAAGCTGCACAT 451  
1099 .....ArgLeuValLysSerIleAsnSerThrGlu 1111  
452 CACACCCCT.....TACACGGCGATTATCATGCGCGTTTG 489  
1111 IalysProValPheThrGluTyrGlnTyrSerGlyHis.....ArgLeu 1125  
490 CATAAATTTGTCACA.....GATGC 509  
1126 GluLysAlaIleAsnAlaLysGlySerGluThrTyrValTyrSerTyrAspAla 1142  
510 AGAA..... 513  
1142 aaSplysLysThrLeuLeuMetThrGlnProAsnGlyArgLysValGlnT 1159  
514 .....CCTGTGAAATGACGAGTACATGAG 540  
1159 YrGlyTyrAsnGluAlaGlyAsnProIleGlnValIleAspAspAlaGlu 1175  
541 GGAATACCTATTCCGATAAAGAAATAT..... 570  
1176 GlyLeuLysIleThrThrAsnThrLysTyrGluGlyAsnAsnValValGln 1192  
571 .....CCGAGCGTCCGATCGGCTGACGACACCATATGGC 610  
1192 uasPvalAspProAsnAspValGlyThrGlyLysAlaThrGlnSerTyrG 1209  
611 GTTATGATGATGACAAACACGCGGATTTA...TCTACTCCGGCGCATG 657  
1209 LntyrAspLysAsp.....GlyAsnValThrSerValLysAspAlaTyr 1223  
658 TTTATGGCGGCAATACACATATGACAGCGTGGGAAATATATGCGTAT 707  
1224 .....GlyThrGlnTyrGluTyrAsn...LysAsnAsnAspValThr 1237  
708 TAGTTGAGCGGCGATGTGCCCATGCCACACAC.....TATGGCCCA 751  
1237 rLysMetLysAspThrGluGlyAsnValThrAspIleAlaTyrAspGly 1254  
752 TGCCGATTGACAGTGGCGGACGACGCGT.....TGCCCAATGTT 795  
1254 euAspAlaValSerGluThrAspLysSerGlyLysSerSerAlaAla 1270  
796 ATTATGACAAACAAACAAATTAATGCTGCTCAACGAGTTTACAAAC 845  
1271 ValTyrAspLysTyrGlyAsnGln.....IleGlnSer 1281  
846 CGGCTACCTTATTCGCGACGGAAGAAAGGTTTCCAGCTGATACGCAAG 895  
1281 rSerLysAspLeuSerAlaSerThrAsn.....IleLeuLysAla 1294  
896 ATTGCTTACGATGACATTTACAGAGC..... 924  
1294 spGlySerPheGluAlaGlnLysSerGlyTyrPasnLeuThrAlaSerLys 1310  
925 GATACACATACCGTCTNTTTTGAACCGCGACGTAACGACATTTTTC.. 972  
1311 AsparGargLysIleSerValIleAlaAspLysSerGlyValLeuSerGln 1327  
973 .....TTTACATCCACACACA 988  
1327 ySerLysAlaLeuGluValLeuSerGlnSerThrSerAlaGlyThrAsp 1344  
989 ACGGTACCGGTACGTAACAGAAACAAAGGNTCCATCAACA... 1035  
1344 isGlyTyrSerSerAlaThrGlnThrValGlnLeuGluProAsnThrThr 1360  
1036 .....AAGCTTAAAGTACAGACAGTCCA.....CT 1061  
1361 TyrThrLeuSerGlyLysIleLysThrAspLeuAlaLysSerArgAlaTyr 1377  
1062 GTTTGACGAATCTTTGATGAATAACTGATAAAGACAGTTTAC..... 1104  
1377 rPheAsnIleAspLeuArgAspLysAspGlnLysArgIleGlnTyrPhe 1394  
1105 .....GCGCACGGGCTGTATGATGATCCGTCAAGTTA 1140  
1394 IsAsnGluTyrSerAlaLeuAlaGlyLysAsnAspThrLysAlaG... 1409  
1141 AACACGGTGAACCTTTCTTATGATTTACGGCAAC...GGCAACT 1187  
1410 .....GlnIleThrPheThrThrProAlaAsnAlaGlyLysAla 1422  
1188 CATCTTATCAACACATCAAC.....CANGCGCGGCGCGTT 1225  
1422 aValValTyrMetGluValAspHisLysAspLysAspGlyLysGlyLysAla 1439  
1226 TGATTTTGA.....GGTGAATTTACGGTCTCG... 1254  
1439 LattPheAspGluValGlnLeuGluLysGlyGlnValSerSerSerTyr 1455  
1255 ...CCTGAAACAAACGAACGTGGCAAGCGCGGCGCTTATATCACTGA 1301  
1456 AsnProValGlnAsnSerSerPheThrSerAla..... 1466  
1302 AGACATACCGTTACTTGGAAGTAAACGCGGTGGCA..... 1338  
1467 .....ThrGluAsnThrPasnValSerGlyAlaSerValAspSerGlu 1481  
1339 .....AAGCAGCGCTGTCCAAATCGGCAAAAGCAGCGCTCACGTT 1380  
1481 LuGlyPheAsnAspAspValSerLeuLysAlaAlaArgThrSerAlaSer 1497  
1381 CAAGCCAAAGGGGAAACCAAGCGCTGATCAGCGTGGCGACGTTACAGT 1430  
1498 GlnAlaGly.....SerValThrLysGlnThrVal 1507  
1431 CATTTGGATGACGCGGACGAGATAA.....G 1459  
1507 lValLeuGlyGlnSerAlaAsnAspLysProValTyrLeuThrLeuThrG 1524  
1460 GCAAAACAAACGCTTGTAGTAATCGCTTGNTCAGCGGACGAGGTACG 1509  
1524 LysMetSerLysAlaSerSerVal..... 1531  
1510 GTGCAACTGAATGCCGATATCATGTTCAACCCGCAACACTGTATTGCG 1559  
1532 .....LysPheThrAspGlnLysAspTyr.... 1539  
1560 CTTTGGCGGCGGACGTTGGATTGAACGGCGCATTCGTTTCGTCACC 1609  
1540 .....SerLeuGlnAlaAsn..... 1544  
1610 GTATTCAAATACGATGAAGGGCGGATGTCNCNATCATATATCCACA 1659  
1545 ...ValThrTyrAlaAspGlySerThrGlyLe...TyrAsnAlaLys 1558  
1660 ACAACATTCACCGTTACCATTCACAGGAATGAATATATACACACGCG 1709  
1559 PheProSerGlyThrGlnGluThrPasnArgAlaValValIleProLys 1575  
1710 TGCGTAAATATCAATAGACTTAAT.....TACACCAAGAAA 1747  
1575 sThrLysProIleAsnLysValAspIleSerIleLeuPheGlnLysSerAla 1592  
1748 TTGCTTACACGCGTGGTTGGCGAG..... 1773  
1592 LathrGlyThrValTyrPheAspAspIleArgLeuIleGlnGlySerLeu 1608



```

1774 .....AAGATACGACCAAAAGCAGCGGCGCTC..... 1803
      |||:|||||
1609 LeuThrIysSerThrTyrAspSerAsnGlyAsnTyrValThrLysGlu 1625
      |||:|||||
1804 ....AACCTGTTACGACGCCCGCCGAGAA...GACCGACCCNGCTGC 1846
      |||:|||||
1625 uAspGluLeuGlyTyrValThrSerThrAspTyrAspGluThrGlyLysL 1642
      |||:|||||
1847 TTTCGGCGGACCAAAATTTAAACGCG..... 1872
      |||:|||||
1642 ystThrSerGluThrAspAlaLysGlyLysThrThrTyrThrTyrAsp 1658
      |||:|||||
1873 .....AACATCAGCAAAACAAACGCGCAACTGTTTTT 1904
      |||:|||||
1659 GluAlaAspGluLeuThrAsnMetThrLeuSerAsnGlyThrSerIleLe 1675
      |||:|||||
1905 CAGCGGAGACCGACACCGCGCTCACAATTCATTAGGAGCGGGGCGT 1954
      |||:|||||
1675 u.....HisThrTyrAspLysGluGlyAsnGluValS 1686
      |||:|||||
1955 CAAAATGGAAGTATCCCAAGAGAAATCGTGGGACAAAGACTGG 2004
      |||:|||||
1686 erLysThr.....IleArgAlaGly..... 1692
      |||:|||||
2005 ATCAGCCGCACTTTAAAGCGGAAATTTCCATATTCAGCGCGGAGGC 2054
      |||:|||||
1693 AlaAspGluThrTyrLysPheGlu...TyrAspAlaMetGlyLysLeuVa 1708
      |||:|||||
2055 GGTGATTTCCCGCATGTGCGAAAGTGGAGGCGATGNCATTGAGCA 2104
      |||:|||||
1708 LysThrThrAspProLeuGlyAsnValIleuAlaSerGluTyrAspAla 1725
      |||:|||||
2105 ATCAGCGCCCAAGCAGTTTGTGTGTGCGACCGCATCAAGCCATACATC 2154
      |||:|||||
1725 sn..... 1725
      |||:|||||
2155 TGACACGTTGCGACGTGACNGCTGCACAATGTGTGCAANAANCAT 2204
      |||:|||||
1726 .....SerAsnLeuThrLysThrIleSerPro..... 1734
      |||:|||||
2205 TACGAGCATAAAGTATGCTTCATTTGACTAAGACNGACNNTAAGCGCA 2254
      |||:|||||
1735 AsnGlyAsnGluValSerLeuSerTyrAspGlyThrAspArg..... 1748
      |||:|||||
2255 NTGTAACNTNCCNATNACGNTNNTNAANCTCNCNGCGCTGCNCA 2304
      |||:|||||
1748 ..... 1748
      |||:|||||
2305 CTAAAGGCAATCTTAGTCAATGCGGATACAGCTTATACAGTCAGCA 2354
      |||:|||||
1749 ...ValLysSerLysSerTyrAsnGlyThrGluLysThrIlePheThrTy 1764
      |||:|||||
2355 CAAGCGCACCCAAACGGCAAC...CTTAGCCCTGCGGCAATGCCCAAG 2401
      |||:|||||
1764 rAsp....LysAsnGlyAsnGluThrSerValIleAsnLysGluGln 1779
      |||:|||||
2402 CAACATTTAATCAAGCCACATTAAGCGCAACATCGGTTGCGGCAAT 2451
      |||:|||||
1779 snThrThrLysLysArgThrPheAspAsnLysAsn..... 1790
      |||:|||||
2452 GCTTCATTATCTAAGCAACAAGCGCGCAAAACGCGAGTGCAGCT 2501
      |||:|||||
1791 .....ArgLeuThrGluLeuThrAspArgGlyLysSerGlnThr 1804
      |||:|||||
2502 T.....TCGACAGACGCTAAGCA..... 2520
      |||:|||||
1804 pThrTyrProSerAspSerAspLysLysThrPheSerTyrIleHisG 1821
      |||:|||||
2521 .....AACGTAAAGCATTCGACATCCAGCATCAGCGCAAT 2550
      |||:|||||
1821 LysAspGluLysGlyThrAsnGlnPheThrTyrAsnLysLeuAspGlnMet 1837
      |||:|||||
2551 GTCTCCCTACCGCATAG...GCAGTATTCATTTT.....GAAAA 2588

```

```

1838 ILeGluMetLysAspSerThrSerSerTyrSerPheAspTyrAspGluAs 1854
      |||:|||||
2589 CAGCCGC.....TTTACCGGACAACTCAGCGGACGAGCAAGANA 2626
      |||:|||||
1854 nGlyAsnValGluThrPheIleThrGlyAsnGlyLysGlyThrSerPheS 1871
      |||:|||||
2627 CAGCATTAACCTTAAAGACAGCAATGACGCTCCGCTCAGCGACGGA 2676
      |||:|||||
1871 erTyr.....AspGluArgAsnLeuValSerSerLeuHis 1882
      |||:|||||
2677 TTAGCAATTTAAACCTTGACAAAGCCACCATTAATCATTCCGCTTA 2726
      |||:|||||
1883 IleGlyAsp.....LysAsnGlyLysAspIleLeuThrGluSerTyr 1896
      |||:|||||
2727 TCGCCACGATGCTGCAGCGCGCCGCAACCGGACGCTGCACACCGCCG 2776
      |||:|||||
1896 rGluTyrAspAlaAsnGly..... 1902
      |||:|||||
2777 GCCGCGGTTGCGCGCTTCATTATTCGTTACACCGCAACTTCGGTA 2826
      |||:|||||
1903 .....AsnArgThrThrIleAsnSerSerAlaSerGlyLysVal 1915
      |||:|||||
2827 GAATCCCGTTTCACACGCTGACGCTAAACGCAATTTGAA..... 2868
      |||:|||||
1916 GlnTyrGluTyr.....GlyLysLeuAsnGlnLeuVa 1926
      |||:|||||
2869 .....NCTCAAGAGACA...TTCCGCTTTATTCGCAACTCT 2902
      |||:|||||
1926 LysGluThrHisGluAspGlyThrValIleGluThrTyrThrGlyLp 1943
      |||:|||||
2903 TCGGCTACGAGAGCAAAATGAGCTGGCGGAAAGTTCCGGAAGNCT 2952
      |||:|||||
1943 heGlyAsnArgLysThrValThrThrIleLysAspGlySerSerLysThr 1959
      |||:|||||
2953 TACACCTTGGCGGTCAACATACCGGCAACGAAACCGTAAGCTCGATCA 3002
      |||:|||||
1960 ValAsnAlaSerPheAsn.....IleMetAsnG 1969
      |||:|||||
3003 ATTTGACGCTAGTCGAAAGGAAAGACAAACACCGCTGCCGAACCTTA 3052
      |||:|||||
1969 nLeuThrLysValAsn.....AspGluSerIleS 1979
      |||:|||||
3053 ATTTCACCTCGCAAAAGCAACAGCTGATGCCGCGCGGTGATCA 3102
      |||:|||||
1979 erTyrAspLysAsnGlyAsnArgThrSerAspGlyLysPheThrTyrThr 1995
      |||:|||||
3103 CTCATCCGCAAGAGCGGAGTTCGCTGCATAATCCGTCGAAGANA 3152
      |||:|||||
1996 TrpAspAlaGluAspAsn...LeuThrAlaValThrLysLysGlyLysAspL 2012
      |||:|||||
3153 AGAGCTTTCGCAACAACCTCGGCAAGGCAAGCAAAACACAGCGGAA 3202
      |||:|||||
2012 yspProPheAlaThrTyrLysTyrAspGluLys...GlyAsnArgIleGln 2027
      |||:|||||
3203 AAGACACGCGCAAAAGCCTTGAGC 3226
      |||:|||||
2028 LysThrValAsnGlyLysValThr 2035
      |||:|||||
seq_name: SwissProt_40:VG37_BPT2
seq_documentation_block:
ID VG37_BPT2 STANDARD; PRT; 1341 AA.
AC P07067;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Long tail fiber protein p37 (Protein Gp37) (Receptor recognizing
DE protein).
GN 37;
OS Bacteriophage T2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like phages.

```



```

388 :||||| 392
389 AlaGlyAspGlyGln.....
1464 AAACAAAGCCTTAGTGAATCGGCTGNTCAGGCG..... 1500
393 .....ThriHisIleGlyTyrAsnSerGlyGlyLysMetSerH 405
1501 .....AGGGTACGGTGCACCTGCAATGCGGATTAATCAG..... 1533
405 IStyrPheArgGlyLysGlyGlnThrAsnIleAsnThrGlnLysGlyMet 421
1534 ...TTCAACCCCGAC.....AAACTCTATTTCGGCTTCGCGGCGGACG 1574
422 GluValAsnProGlyIleLeuLysLeuValThrAspSerAsnValGln 438
1575 TTGGATTAAACGGGCACTTCGCTTCGTCACCGTATTCAAATATCCG 1624
438 nPheTyrAlaAsnGlyThrValSerSerIleGlnArgIleLys...PheA 454
1625 ATGAAGGGCGGATGATGNCNATCATATGACACACACATCCACGGTT 1674
454 spasnGly.....Leu 457
1675 ACCATTACAGGGAATGAAAGTATTACACACACGAGTGTAGAAATATCAA 1724
458 ValLeuThrGly.....AlaArgProAspGlyIleGlnLeuAs 470
1725 TAGACTTATATACAGCAAAAGAAATTCCTACACGGTTGGTGGCAGA 1774
470 P..... 470
1775 AAGATACGACCAAAACGAGCGGGCTCAACCTGTTTACACCGCGCC 1824
471 .....AlaProThr 473
1825 GCAGAGAGC.....CGCACCCGCTGCTTCCGCGGAAACAAATTTAAA 1868
474 AlaAlaAspGlyThrLysThrIleLeuThrAlaGlyLysThrArgAlaGln 490
1869 CGGCAACATCAGCGAAACAGCGCAAACTGTTTTCAGCGGCGACCGGA 1918
490 YGlnAsn.....LysSerTyrValSerIleLysAlaT 501
1919 CACGCAAGCCTACATCATTTAGGAAGCGGGTGTCAAAATAGAGAGT 1968
501 rPGLysAsnSerPheAsn.....AlaSerGlyAspArgAlaArgGlnThr 515
1969 ATCCCAAGAGAGAATCGTGTGGACACGACTGATCNACCGCACGTT 2018
516 Val..... 516
2019 TAAAGCGGAAATTCATATTCAGGCGGCGAGCG.....GTGATTT 2062
517 .....PheGluValGlyAspGlyGlnGlyPheHisPheTyrS 529
2063 CCGCGAATGTGTCCAAAGTGAAGCGATGNCATTTGAGCAATCAGCGC 2112
529 eArgIleAlaValAlaProAlaProGlySerThr..... 539
2113 CAAGCACTTTTGGTGTGCGACCGCATCAAGCCATCAATCTGTACAG 2162
540 .....ValGlyProIleGlnLeuArgValAsn..... 548
2163 TTGGACTGCGAGCNGTCTGACAATTTGTGTGAAANAANCAATTCAGCAG 2212
549 .....GlyGlyLeu.....LeuThrAlaG 555
2213 ATAAAGTATGTTGCTCATTTGACTAGACAGACANTNAGCGCANTGTAGN 2262
555 LysSerIleValAlaSerGlySerIleThrThrGlnSer..... 567
2263 CTNNCCATNATNCGTNNTTNAANCTCNCNGCGCTCNCNCACTNAANG 2312
:::
:::
:::
SerLeuAsnValAlaAsn 573
2313 CAATCTTAGTGCAAATGCGGATACAGCTATATACAGTACCGCACAGCCA 2362
573 nGlyLeuSerValAsnGlyGlnAlaLysPheGlyGlyThrAlaAsnAlaL 590
2363 CCCAAAGCGCACACTTAGC.....CTCGTGGCGAATGCCCAAGCA 2403
590 euArgIleThrPasnAlaGluTyrGlyValIlePheArgArgSerGlnSer 606
2404 ACATTTAATCAACGCCACATTAAACGGCAACNCATCGGNTTCGGCAATGC 2453
607 AsnPheTyrIleIleProThrAsnGlnAsnGlnGlyLysSerGlyAspL 623
2454 TTCATTATCTAAGC.....AACAAAGCCCGCA... 2481
623 eHisSerSerLeuArgProValArgIleGlyLeuAsnAspGlyAlaValG 640
2482 .....CAAAAGCGGAGTCTG...ACG 2499
640 LysLeuGlyArgAspSerPheIleValAspGlnAsnAlaLeuThrThr 656
2500 CTTTCGACACACGCTAAGCGCAACGTAAGCATTCCGCACTCAACGGCAA 2549
657 IleAsnSerAsnSerArgIleAsnAlaAsnPheArgMetGlnLeuGlyG 673
2550 TGTCTCCCTAAGCCGAT..... 2565
673 nSerThrTyrIleAspAlaGluCysThrAspThrValArgProAlaGly 690
2566 .....AAGCAGATATTCAT 2580
690 LysIleSerPheValSerGlnAsnAsnGlnAsnValArgAlaProPheTyr 706
2581 TTGAA..... 2586
707 MetAsnIleAsnArgThrAspThrSerThrTyrValProIleLeuLysG 723
2587 .....AACAGCCGCTTTACG...GGACACTCTACGGCGCA 2617
723 nArgTyrValGlnGlyAsnSerCysTyrSerLeuGlyThrLeuIleSer 740
2618 GCAAGGANNACGATACACTTA..... 2640
740 hreLysAspPheArgIleHisTyrHisGlyGlyLysAspAsnLysSerH 756
2641 .....AAAGACAGCGAATG 2654
757 GlyProGlnLysAlaAspLeuAlaTrpGlnPheArgArgAspGlySerH 773
2655 GACGTCGCGTCAAGCAGCAATTA.....G 2680
773 eArgSerProAsnLysIleGluIleAsnAlaValThrIleGlyThrAspG 790
2681 GCAATTTA.....AACCTGACACAGCCGACATTCACATC 2715
790 LysAsnIleThrGlyGlyThrGlyAsnPheAlaAsnLeuAsnThrThrLeu 806
2716 AATTCGCGCTATCGCACGATGCTGAGCGGCGCAACCGCGCAGNGTGC 2765
807 Asn.....ArgLysThrThrValGlyGlyThrPalagLysSer..... 818
2766 AGACAGCGCGCGCGCGCTTCGCGCGCTTCCTATTA..... 2802
819 .....SerValValGlyTyrTyrLysP 826
2803 .....TCCGTTACACGCGCACTGCGGTAGATCCGTTTCAACAGCGG 2847
826 hreAlaThrValThrIleProGlnSer.....ThrGly 836
2848 ACGGTAAACGCAAAATTAAGACNGTCAAGGACATTCGCTTATGTGGA 2897
837 ThrValThrPheLysIleSerGlyGlyAlaGlyPheAsnPheLysSer... 852

```

```

2898 ACTCTGCGCTACCGAAGCAGCAATTTGAGCTGGCGAAGATTCCGAG 2947
853 .....TyrAsnGlnAlaSerIleAlaGlu..... 860
2948 GNACTTACACCTGGCGGTCAACATACCGCAACGAGACC..... 2988
861 .....IleValIleuArgThrGlyAsnAsnProIleGlyIle 872
2989 .....GTAGCTCGATCAATGACGGT 3011
873 AsnAlaValIleuTrpAsnArgSerAspLeuSerPheAsnGlnIleAlaThr 889
3012 AGTGGAGGAGAAAGACAGACAAA..... 3033
889 rMetAsnThrSerAspThrThrAspValTyrValPheCysGluGlyT 906
3034 .....CCGCTGCGCAAAACCTTAATTTC 3057
906 yThrAsnAlaLeuIleValGluTyrSerCysSerGluAsnSerSerVal 922
3058 ACCCTGCAAAAGCAACAC.....GTGCGATGGCGGGCGC 3089
923 ThrValValGlyLeuAsnGlyGlyIleGlnProValValAspAla.... 937
3090 GTGGCGTTACCACTCATCCGCAAGACGCGGAGTTCGCGCTGCATATC 3139
938 .....LeuProGluGly.....HisValVal 944
3140 CGGTCAAGAACAGAGCTTCCGACAAATCGCGCAGCAAGCCAA 3189
944 aGlyIlySThrValArgLeuLeuAsnIleGlyGlyThrPheAlaAla 960
3190 AAACAGGCGGGA..... 3201
961 GlyIleuSerAspIleValThrArgGlyGluTyrValAlaAspAsnGlnIly 977
3202 .....AAAGACAACGCGCAAGCCTTGACGCG..... 3228
977 sclyMetArgIleuSerAsnGlyAsnAsnIleGlySerAsnAlaAlaI 994
3229 .....CTGANTGCGCGCGCGCGCAT 3249
994 IeLeuArgAsnAspIlyGlySerPheTyrIleLeuAlaThrAspLysAsn 1010
3250 GCGCGCGCAAAAGACAGCAAGC..... 3270
1011 ThrAlaGluLysSerAspAlaAlaAsnGlyAspTrpAsnGlyLeuArgPr 1027
3271 .....GTTGCCGAAACGCGCGCGCGGAGGCGGCAAAATCTCGCATTA 3315
1027 oPheAlaIleAsnMetAlaAspGlyArgValGlyMetAsnHisGlyLeuA 1044
3316 T...GCAGCGGAGGAGAGAGAAAAACGGGTGCAGCGCGGATTAAGACAG 3352
1044 snIleThrGlyGlyLeuAsnValThrGly..... 1054
3353 CGCNTTGGCGAAACAGCGGCAAGGGAACCGCGCGGTACACCGGCT 3412
1054 ..... 1054
3413 TCCTCCGCGCGCGCGCGCGCGGATTTGCGCAACCGCAGCCCA 3462
1055 .....GlyLeuAsnValThrSerGlyAsnT 1063
3463 CCGCACTCAACCCCAACCGCAGCGCATGATNAGCCCTTATGCCAA 3512
1063 nIleSerGlyAsnIleSerSerArgValValAlaArgTTPArgGlyAl 1079
3513 TAGCGGTGAGTGAATTTCCGCGCAGCTCAACAGCTTTCCGCGGAC 3552
1079 aSerGlyTTPAlaAspAsnSerAspThrMetLysSerLysIleThrPheM 1096

```

```

3563 AGGACGAATTTGACCGCGTGTGGCCGAGACGCCGCAACCGCTGTGG 3612
1096 eTrAlaSpHISGlyAspLeuSerAsnSerAsp..... 1106
3613 ACAGCNGCATCCGACACCAACACTACCTTGCAGATTTCCGCCG 3662
1107 ...SerTyrTyrProIleValGlyAlaTyrSerAsnTyrGlySerAlaG 1122
3663 CTACGCGCAAAACAGCAGCTGCGGCAAAATCGTATGCAAGAAACCTG 3712
1122 yTyrArgGlnThrPheGluPheIleTyrValGlySerGlyThrThrAlaG 1139
3713 GCAGCGCGCGCGTTCGATCTGTTTCCGACAAACCGGACCGAAGAAC 3762
1139 Ly...TrpArgAspGlyIleIleArgIleArgGly..... 1149
3763 TTGACAGCGGCGATCGCACTCGGACGCGCTTGGCCAGCGCGCTTT 3812
1150 ...AspAsnAlaAsnGlyGlnGlnAlaArgTTPArg.....Ph 1161
3813 CCGCAATACGCG.....A 3826
1161 eThrMetAspGlyThrLeuAspCysProGlyLysValIleLeuLeuProGlnT 1178
3827 TCGCAGAGTTTCGACATCGGCATCAGCAGCGCGCGGTTTTCAGACGCG 3876
1178 hrcIyAlaPheGlyValAsnThrSerAsnGlyLeuGlyLysAsnSerIle 1194
3877 AATCTTTCAGAGCGCATCGAGGCAAAATCCGCGCGCGCTGCTGATTA 3926
1195 ThrPheGlyAspSerAspThrGly...IleLysGlnAsn..... 1206
3927 CGGCATTCAGCAGCATACCGCGCGCTTTCGCGGATTCGCGCANTGAC 3976
1207 .....GlyAspGlyLeuLeuAspI 1213
3977 CGTACATCGCGCGCACCGCTATTC...GTCCAAAACCGGAT...TAC 4020
1213 IeTyrAlaAsnSerValGlnValPheArgPheGlnAsnGlyLysLeuTyr 1229
4021 CGCTACGAAACGTCATATTCGCCACCGCGGCTGCTTCACCGGTA 4070
1230 SerTyrLysAsnIleAsn.....AlaProAsnValIly 1240
4071 C...CNGCGCGCATTAAGCAGATTTATTCATCAACCGCGCAACAC 4116
1240 rIleArgSerAspIleArgLeuLysSerAsnPheLysProIleGluAsn 1256

```

seq\_name: SwissProt\_40:VGLX\_HSVEB

```

seq_documentation_block:
ID VGLX_HSVEB STANDARD: PRT; 797 AA.
AC P28968:
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Glycoprotein X precursor.
GN 71.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1."
RL Virology 189:304-316(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina: Saccharomycetes;



710 PheThr.....SerAsnSerThrGluThrIle.....AlaVal 720  
2220 GATTGCTTCATTGACTAGACNGACNTNAGCGGCAATGTNAGNCTNNCN 2269  
|||:|||||  
720 IleSerAsnIleHisLysThr..... 727  
2270 ATNCGNNTNTNMAANCTCNCNGCGCTGCNCACTNANGCAATCTT 2319  
727 ..... 727  
2320 AGTGCAAATGGCGATACACGTTTATACATGCACACCAACGCCACCAAAA 2369  
|||:|||||  
2328 SerSerAsnLysAspTyrSerLeuThrThrThrGlnIleuLysThr...Se 743  
CGGCAACCTTAGCTCGCTCGTGGGCAATGCCCAAGCAACATTTAATCAACCA 2419  
|||:|||||  
743 rGlyLysGlnThrLeuValLeuSerThrValThrThr.....T 756  
2420 CATTAACGGCGACNCAATCGGNTTCGGGCAAT.....GCTCATTT 2460  
|||:|||||  
756 hrValAsnGlyAlaIleAlaThrGluTyrThrThrTyrCysProAlaSerSer 772  
2461 AATGTAAAGCAACAAGCCGCAACAAAGCAATGTCAGCTTTCGACAA 2510  
|||:|||||  
773 IleAlaTyrThrThrSerIleSerTyrLysThrLeuValLeuThrThrGln 789  
2511 CGGTAAAGCAACGATAGCCATTCGCGACATCCGCAACGGCAATGTCTCCCTAG 2560  
|||:|||||  
789 uVal.....CysSerHisSerGluCysThrProThrVal..... 800  
2561 CCGATTAAAGCAATATTCATTTGTGAAAAAGCCGCTTTACCGGACAATCTC 2610  
|||:|||||  
801 .....IleThrSerValThr 805  
2611 AGCGGACAGCAAGANACAGATTACACTTAAGACAGCAATGGACGCT 2660  
|||:|||||  
806 AlaThrSerSerThrIleProLeuLeuSerThrSerSerThrValIle 822  
2661 GCCGTGAGGACGGAATTAAGCAATTTAAACCTTTGACAAAGCCACCATTA 2710  
|||:|||||  
822 uSerSerThrValSerGlnLysAlaLysAsnProAlaIleSerGluValT 839  
2711 CACTCAATTCGGCTATCGCGACGATGCTGACGGCGCAACCGGCAAGN 2760  
|||:|||||  
839 hrIleAsnThr.....Gln 843  
2761 GTGTCAACACAGCCGCGCCGCTTCGCGCGCTTCCCTATTATCCGTTAC 2810  
|||:|||||  
844 ValSerIleAlaThrSerGlnAlaThrSerThrSerThrGlnValSerAlaThr 860  
2811 ACCGCAACCTTCGGTGAATCCCGTTTCAACAGCTGACGGTAAAGGCA 2860  
|||:|||||  
860 rSerAlaThrAlaThrAlaThrAlaSerGluSerSerThrThrSer..... 873  
2861 AATTGAACNCTCAAGAACATTCGCTTATATGCGGAATCTTTCGGCTAC 2910  
787 ..... 873  
2911 CGAAGGACAAATTAAGACTGGCGGAAGCTCCGAGNACTTACACTT 2960  
|||:|||||  
874 .....GlnValSerThrAlaSerGluThrIleSerThrIle 885  
2961 GGGGATCAACAAT.....ACCGGCAAC.....GAACCCGTTA 2992  
|||:|||||  
885 uGlyThrGlnAsnThrThrThrThrThrLysLeuLeuPheProAlaLeuS 902  
2993 GCGTCGATCAATTC.....ACGGTATGGAAGGGAAGAACAA 3031  
|||:|||||  
902 eThrGlnMetIleAsnThrThrValVal.SerArgLysThrLeuIleI 918  
3032 .....AACCGCTGTCCGAAACCTTAATTTAC 3059  
918 eSerThrGlnValCysSerHisSerLysCysValProThrValIleThrI 935

```

3824 .... GCATCGGCGAGGTTCGACATCGCATCGACGCGCGGCTTTTA 3868
CC .....
1180 hrLeuLysAlaThrThrGlnThrSerSerThrAlaLysValArgIleThr 1196
3869 GCAGCGCGCATCTNT.....CAGACGGCATCGGA 3897
CC .....
1197 SerLysSerLysAlaThrSerThrSerIleSerLeuSerThrSerThrGln 1213
3898 GCGAAATCGCGCGCC.....GCGTCGTCGATTACCG 3929
1213 uSerGlnSerSerSerGlyTyrLeuSerLysGlyValGlySerGlyThrG 1230
3930 CATTCAGCGCGCAT.....ACCGCGCGGCTTCG..... 3958
1230 LucYsthrGlnAspValProThrGlnSerSerSerProLysSerThrLeu 1246
3959 .....GCGCATTCGCGCATCGACCGTACATCGCGCAACGCGCATTTTC 4002
1247 AlaTyrSerProSerValSerThrSerSerSerSerSerSerThrThr 1263
4003 GTCCAAAAGCGATTACGCGCTACGAAACGTATATGCGCACCCCGG 4052
1263 rThrAlaSerThrLeuThrSerThrIsthrSerVal.....ProL 1277
4053 TCCTGCGTTCACCGATACCGCGCGCATTAAGCAGATTATTCATTCA 4102
1277 euLeuProSerSerSerSerIleSerAlaSerSerProSerSerThrSer 1293
4103 AACCGGCGCAACACATNTCCATCAGCNCCTTATTNAGCTGTCCTATACC 4152
1294 LeuLeuSerThrSerLeuProSerProAlaPheThrSerSerThrLeuPr 1310
4153 G.....ATGCCCG 4160
1310 oThrAlaThrAlaValSerSerSerThrPheIleAlaSerSerLeuProL 1327
4161 TTGCGGCAAGTCCGACACGCGTCATATACCG 4192
1327 euSerSerLysSerSerLeuSerLeuSerPro 1337
seq_name: SwissProt_40:HLA_SERMA
seq_documentation_block:
ID HLYA_SERMA STANDARD; PRT; 1608 AA.
AC P15320;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Hemolysin precursor.
GN SHLA.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.
RC STRAIN=SN8;
RX MEDLINE=8257037; PubMed=3290200;
RA Poole K., Schiele E., Braun V.;
RT "Molecular characterization of the hemolysin determinant of Serratia
marcescens."
RL J. Bacteriol. 170:3177-3188(1988).
CC -1- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
DEFINED.
CC -1- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA
REQUIRES SHLB FUNCTION.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: TO P. MIRABILIS HEMOLYSIN (HPMA).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL; M22618; AAA50323.1; -
DR PIR; A28182; A28182.
KW Hemolysin; Toxin; Outer membrane; Signal.
FT STGMAL 1 30
FT CHAIN 31 1608 HEMOLYSIN.
SQ SEQUENCE 1608 AA; 165078 MW; D669B476F7DAD51 CRC64;

alignment_scores:
Quality: 182.50 Length: 1054
Ratio: 0.369 Gaps: 56
Percent Similarity: 46.964 Percent Identity: 21.157

alignment_block:
US-09-303-518D-651 x HLYA_SERMA ..
Align seg 1/1 to: HLYA_SERMA from: 1 to: 1608

757 ATTGCGAGTGGCGGACGCGACGCGGTTGCCAATGTTTATTATTCAGCA 806
::: |||||: |||||: ::::
657 ValIleGlySerGlnValAlaLysSerGlyGlyLeuSerValGlySerLys 673
807 AACAAACAATTAATGCCCTGCTCAAC..... 831
||||: |||||: |||||: |||||:
673 ThrIleLysN.....IleAsnValLysAlaLysGlnGlnGln 687
832 .....GGAGTTTACAACCGCGCTACCTTATTC 861
687 snIleAspGlnGlnLysThrAlaLeuThrValAsnGlyTyr..... 700
862 GCGACGAGAAACGCTTTCAGCTGATACGCAAGATTGCTCTAGATCA 911
::: |||||: |||||: |||||: |||||:
701 AlaLysGlnAlaGly.....AspLys 707
912 CATTCACAGAGCGCAT.....ACACATACCGCTCTNTTTGAACCGC 952
|||||: |||||: |||||: |||||:
707 sGlnTyrArgAlaGlyLeuArgIleGlnIsthrArgAspSerGlyLysT 724
953 GCAGTACGCGACATTTTCCTTTACATCCACACACGATGAGGTACG 1002
::: |||||: |||||: |||||: |||||:
724 hrThrArgThrGlnAsnSerLysSerLeuSerGlySerValLys 740
1003 GTA..ACGAACAACAAGAAAGTTCCAATCCAAAGCTT..... 1041
::: |||||: |||||: |||||: |||||:
741 LeuLysAlaGlnLysAspValThrPheSerGlySerLysLeuValAlaAs 757
1042 .....AAAGTACGACAGTCCGACGT 1063
757 pLysGlyAspAlaSerValSerGlyAsnLysValSerPheLeuAlaAla 774
1064 TTGACGATCTTTGATGAATGAATGAATGAATGAATGAATGAATGAAT 1101
|||||: |||||: |||||: |||||: |||||:
774 spAspLysThrAlaSerAsnThrGlnGlnThrLysIleGlyGlyLysPhe 790
1102 TACGCGCAGGGGGTGTATATCAGTACCGTCCAGATTAACACGATGCA 1151
||| ::|||: |||||: |||||: |||||: |||||:
791 TyrTyrThrGlyGlyLysPhe.....LysLeuGlySerGlyLys 803
1152 AACCTTCTTTTATGATTACGACGCGCAACTCATTTATCAACA 1201
::: |||||: |||||: |||||: |||||:
803 IGIuAlaGlyTyr.....GluAsp 810
1202 ACATCACCAAGC.....GCGGCGGTTTGAATTTT 1233
|| ::|||: |||||: |||||: |||||:
810 snLysThrGlnAlaGlnSerLysAlaIleThrSerLysSerAspVal 826
1234 GAAGGTGATTTTACGCTCGCGTCAAAACAAGAAACGATGCAAGC.. 1281

```



827 LysLysAsnLeuThrIleAsnAlaArgAspLysLeuThrGlnGlnGlyAla 843  
1282 ..... GCGGGGCT... C 1291  
843 aglnHisSerValGlyGlyAlaTyrGlnGlnAsnAlaAlaGlyValAsp 860  
1292 ATATCACT..... GAAGACGATACCGTTACTTGAAA..... 1323  
860 IsIeuAlaAlaAlaAspThrAlaSerThrThrThrLysThrAspVal 876  
1324 ... GTAAAGCGCGTGGCAACGACCGCTTCC..... 1353  
877 GlyValAsnIleGlyAlaAsnValAspTyrSerAlaValThrArgProVa 893  
1354 ..... AAATCGGCAAGGACGCGTGCACGTTCAAGCCAAAGGGGAAA 1396  
893 IGIuArgAlaValAlaGlyLysAla..... AlaLysLeuAspA 905  
1397 ACCAAGCGCTGCATCAGC..... GTGGCGCAGCGGTACAGTCATT 1434  
905 IatThrGlyValIleAsnAspIleGlyGlyTyrIleGlyAlaProAsnValGly 921  
1435 TTGGAT..... CAGCAGCGACGATTAAGSCAAAAACACCTTTAG 1478  
922 LeuAspIleGlyAlaGlnGlyLysSerSerGluLysArgSerSerSerSe 938  
1479 TGAATCGCTGTGTCAGCGCGCAG... GGTACGTCGACATGAATGCGC 1525  
938 rGlnAlaValAlaSerSerValGlnAlaGlySerIleAspIleAsnAla 955  
1526 ATATCACTTCAACCCGCAACACTATTTCGCTTCGCGCGCAGCGT 1575  
955 yseGlyValAlaArgAspGlnGlyThrGlnTyrGlnAlaSerLysGlyAla 971  
1576 TTGGATTAAAGCGGCATTCGCTTCGCTTCACCGGTAATTCAAATACCGA 1625  
972 ValAsnLeuThrAlaAspSerHisArgSerGluAlaAlaAlaAsnArgI 988  
1626 TGAAGGCGCGATGATGTCNCATATATGCGCACACACATCCACCGTTA 1675  
988 nAspGlnGlnSerIArgAspThrArgGlySerAlaGlyValArgValTyrT 1005  
1676 CCATTACAGGGAATGAAGTATTACACACCGAGTGTAG... AATATC 1722  
1005 hrThrThrGlySerAspLeuThrValAspAlaLysGlyGlnGlyThr 1021  
1723 AATAGACTTAATTACAGCAAGAAATTCCTACACGCTTGTTGGCGA 1772  
1022 GlnArgSerSerSerAlaSerGlnAlaValThrClySerIleAspAl 1038  
1773 GAAGATACGACCAAAACGAGCGCGGCTCACCTTGTTACACCGCG 1822  
1038 aAlaAsnGlyIleAsnValAsnValLysLysAspAlaIleTyrGlnGlyT 1055  
1823 CCGCAGAAGACGCGACCCGCTCTTCGCGCGA..... 1857  
1055 hrAla..... LeuAsnGlyLysArgGlyLysThrAla 1065  
1858 ACAATTTAAACGGCAACATC..... 1878  
1066 ValAsnAlaGlyLysPleArgLeuAspGlnAlaSerAspLysGlnSe 1082  
1879 ..... ACGCAAAACAAACGCAACGTTTTCAGCGGAGACCGA 1918  
1082 rGluSerArgSerGlyPheAsnValLysAlaSerAlaLysGlyGlyPheT 1099  
1919 CACCGCAGCGCTTACATTTAGAAAGCGGTGTCAAAATAGAGAGT 1968  
1099 hrAlaAspSerLysAsn... PheGlyAlaGlyPheGly..... GlyGly 1112  
1969 ATCCCAAGAGCAAAATCGTGTGGCAACGACTGATCNACCGCAGCTT 2018  
:::|||||

1113 ThrHisAsnGlyLu..... SerSerSer 1121  
2019 TAAAGCGGAATTTCCATATTCAGGGGAGCGGCTG..... 2058  
1121 ThrAlaGlnValGlyAsnIleSerGlyGlnGlnGlyValGluLeuLysA 1138  
2059 ..... ATTCCCGCAATGTTGCCAAAGTGAAGCGCAT 2091  
1138 IagLysArgAspLeuThrIleuGlnGlyThrAspValLysSerGlnGlyAsp 1154  
2092 TGCNATTTGACCAATCAGCCCGACGACTTTTGTGTGCGACCGCATCA 2141  
1155 ValSerLeuSer..... AlaGlyAsnLysValAlaLeuGlnAl 1167  
2142 AAGCCATCAATCTGTACAGCTGTGGACTGACGACNGCTGTACAAATTTGTG 2191  
1167 aAlaGluSerThrGlnThrArgLysGlu..... 1176  
2192 TCGAANNAANCAATTACCGACGATTAAGTGAATGCTTCACTGATGAACN 2241  
1177 ..... Ser 1177  
2242 GACNTNAGCGGCAATGTNAGCTNNCCNATNACGTTNNTTNANCTGNC 2291  
1178 LysLeuSerGlyAsnIleAspLeuGlyAlaGlySer..... As 1191  
2292 NGCGCTGCGNCCATNANAGCAATCTTAGTGCAAAATGCGATACAGTT 2341  
1191 pSerLysGlyLysThrGlyGlyLysAsnLeuSerAlaGlyLysAlaPheAspI 1208  
2342 ATACA... GTCAGCCACAAGCGCCACCAAAAGCGCACTTACGCTCGT 2388  
1208 IeAlaLysValAsnGluSerAlaThrGluArgGlnGlyAlaThrIleAla 1224  
2389 GGCATGCGCCAAAGCAACATTT..... AATCAAGC 2417  
1225 SerAspGlyLysValThrLeuSerAlaAsnGlyLysGlyAspAspAlaLe 1241  
2418 CACATTTAAAGCGCAACNATCGGNTTCGGCAATGCTTCACTTATATCTAA 2467  
1241 whIstLeuGlnGlyAlaLysValSerGlyLysLeuAlaLeuGlu... 1256  
2468 GCACACAGCGCGCAAAACGCGACTGTGAGCTTTCGACACGCTTAAG 2517  
1257 ..... AlalysAsnGlyLysIleLeuLeu... GluSerAlaLys 1268  
2518 GCAAACGTAAGCCATTCGCGACTCAAGCGCAATGCTCTCCATGACGATTA 2567  
1269 ... AsnGlnGlnHis..... LysAspAsnThrPserLeuGlyIleLys 1281  
2568 GGCAGTATTCATTTTGAACACGCGGCTTACCGGACACATCGAGCGCA 2617  
1281 sAla..... AsnAlaLysGlyGlnThrPheAsnLysAspAlaGlyG 1296  
2618 GCAGAGANACACCATTTACACTTAAAGACAGAGATGAGCGTCCGTCA 2667  
1296 LysLysValAspProAsnThrGlyLysAspThrHis... ThrLeuGlyAla 1311  
2668 GGCACGGAATTAGC..... AATT 2687  
1312 GlyLeuLysValGlyValGlnGlnGlnAspLysThrThrHisAlaAsnTh 1328  
2688 AAACCTTGACAAAGCCACCATTTACATCAATTCGCGCTATGCGACGATG 2727  
1328 rGlyLysThrAlaGlyAspValThrLeuAsnSerGlyLysAspThrArgL 1345  
2738 CTGACAGCGCGCAAC... GGCAGNGTGTCAAGC 2769  
1345 eAlaGlyAlaArgValAspAlaAspSerValGlnGlyLysValGlyGly 1361  
2770 ACGCGCGCGCGCTGCGCGCGCTTCCTTA..... 2799  
1362 AspLeuHisValGluSerArgLysAspValGluAsnGlyValLysValAs 1378  
:::|||||

```

2800 .....TTATCGGTTACACCGCCAACTTCGGTAGAT 2830
1378 pvalaspalaaglyleuserhisserasnasprioglyserthlethr 1395
2831 CCCGTTTCAC.....ACGCTGACGGTAAACGCCAAATTAAGANGT 2871
1395 erlyleuserlyvalglthrproargtyrvalaglyvallysglu 1411
2872 CAAGAACATTCGCTTATGTCGGAACCTTCGGCTACCGAACGCGAA 2921
1412 .lyseuaduvalaglyval.....AsnlysvaialaspalaThr 1425
2922 ATTGAAGTCGGGAAATCTCCGAGNACTTACACCTTCGGCGTCA... 2968
1425 hraspLyTyrAsnSerValAlaIargArgLeuAspProGlnGlnAspThr 1441
2969 .....ACAAATACCGGC 2979
1442 ThrGlyAlaValSerPheSerlyAlaGluGlyValThrLeuProAl 1458
2980 AAGC.....AACCGTAAGCTCGATCAATTGACGGTAGTGA 3017
1458 atHproAlaglygluysPro.....G 1466
3018 AGGGAAGACAAACACCGCTGCCGAAACCTTAATTCACCTCGCAA 3067
1466 lnglyProleuTyrAspArgGlyAlaIargThrValGlyGlyAlaValLys 1482
3068 ACGAACACGTGATGCCGCG.....CG 3090
1483 AspSerIleThrGlyProAlaGlyArgGlnGlnHisLeuysValAsnAl 1499
3091 TGGCGTTACCACTCATCCGCAAGACGGCGAGTCCGCTGCAATATCC 3140
1499 aspyAlaValAsnAsnAlaValGlyGluGlnSerAla.....IleA 1514
3141 GGTCAAGACAGAGCTTCCGCAAACTCGCAGACGACAGCAAAA 3190
1514 laglyLysAsn.GlyValAlaLeuGlnVal...GlyGlyGlnThrGlnLe 1529
3191 AACAGCGCGGAAAGACACCGCGCAACCTTGACCGCTATTCGGCC 3240
1529 urthelyGlyGlu...IleArgSerGlnGlnGlyValGlnLeuGlnLe 1545
3241 GGGCGCGATGCCCGCAAGACAGCAAGCGCTGCCGACCGCGCGCN 3290
1545 lySerGlnValSerGlnGlnAspValAsn..... 1554
3291 GGCAGCGCGGGAATAATGCGCATTAATGACGCGAGAGAGAAAC 3340
1555 GlyIlnArg.....TyrGlnGlyGlyArgValAlaSPAl 1566
3341 GGGTGCA.....GGCGATTAAGACAGCGCNTTGGCGAAACACGCGAA 3384
1566 aalaIalThrValGlyGlyLeuGlyGlyAlaAlaValGlnSerValA 1583
3385 GCGGAAC 3392
1583 laglyAsn 1585
seq_name: SwissProt_40:P3P_LACLC

```

```

OG Plasmid.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.
RC STRAIN=SK11;
RX MEDLINE=89340435; PubMed=2760036;
RA Vos P., Simons G., Siezen R.J., de Vos W.M.;
RT "Primary structure and organization of the gene for a procaryotic,
RT cell envelope-located serine proteinase.";
RL J. Biol. Chem. 264:13579-13585(1989).
CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -1- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC specificity, although some substrate preference have been noted,
CC e.g. large hydrophobic residues in the P1 and P4 positions, and
CC Pro in the P2 position. Best known for its action on caseins,
CC although it has been shown to hydrolyse hemoglobin and oxidized
CC insulin B-chain.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J04962; AAA0333.1; ALT_SEQ.
DR PIR: A32634; A32634.
DR HSRP: P00782; 2SPT.
DR MEROPS: S08.019; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR003137; PA.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR Pfam: PF02225; PA; 1.
DR Pfam: PF00082; Peptidase_S8; 3.
DR PRINTS: PR00723; SUBTILISTM.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
KM Hydrolyse: Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KM Transmembrane.
KW SIGNAL
FT PROPEP 1 33
FT CHAIN 34 187
FT DOMAIN 188 1902
FT TRANSMEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT DOMAIN 1867 1872
SQ SEQUENCE 1902 AA; 200550 MW; 87CECBA9345F9D3 CRC64;

```

```

alignment_scores:
  Quality: 181.50      Length: 1558
  Ratio: 0.265         Gaps: 74
  Percent Similarity: 44.031  Percent Identity: 18.678

```

```

alignment_block:
  US-09-303-518d-651 x P3P_LACLC  ..
  Align seg 1/1 to: P3P_LACLC from: 1 to: 1902

```

```
124 GCGGGACACATTATTTGGCATCAACTACCAATACCTATCGGCGACTTGC 173
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
387 SerGlyThrGlnGlyValAsnLysAspTyr..... 399
174 CGAAATTAAGGCAAGTTTCAGTCCGGGGGGAAGATATTGAGCTTACA 223
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
400 .....GlyLeuGlnAspAsnGluMetValG 408
224 ACAAAAAGGAGAGTTGGTGGC.....AATCAATGACAAA 261
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
408 LysErProGlyThrSerArgGlyAlaThrThrValAlaSerAlaGluAsn 424
262 GCGCCGATGATGATTTTCTGTGTGTGGCGGTAACGGCGGCGGCGCAAT 311
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
425 ThrAspAlaIleThrGlnAlaValThrIleThrAspGlyThrGlyLeuG 441
312 GGTGGGCGATCATATATGTGTGAGCGTGACATAC.....GGGGCT 355
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
441 nLeuGlyProGlyThrIleGlnLeuSerSerHisAspPheThrGlySer 458
356 ATAAC.....AACGTTGAT 369
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
458 heAspGlnLysLysPheTyrIleValLysAspAlaSerGlyAsnLeuSer 474
370 TTTGGTGGGAGGAGAAATCCGATCAGCACCGGTTTTCTTACCAAT 419
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
475 LysGlyAlaLeuAlaAspTyrThrAlaAspAlaLysGlyLysIleAla 491
420 TGTGAAAAGAAATATATTATAGCCTGACCAATTCACACCTTACACGGG 469
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 eValLysArgGlyLupheserPheAspLysGlnLysTyrAlaGlnA 508
470 ATTANCATATGCCCGTTTGCATAAATTTGTCCAGATGAGACACTGTC 519
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
508 IalaGlyAlaIalaGlyLeuIleIleValAsnThrAspGlyThrAlaThr 524
520 GAATGACGATGATGAGGGGAATACCTATTCC..... 555
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
525 ProMetThrSerIleAlaLeuThrThrThrPheProThrPheGlyLeuSe 541
556 .....GATAAAGAAAATATCCGACGCTGCCCATCGGCTCAGCAGC 598
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
541 rSerValThrGlyGlnLysLeuValAspTyrVal.....ThrAlaH 555
599 AC.....CACTAT 606
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
555 isProAspSerLeuGlyValLysIleThrLeuAlaMetLeuProAsn 571
607 TGCGGTATGATGATGACAAACAGCGGATTTATCTACTCCGCG..... 651
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
572 GlnLysTyrThrGlnAspLysMetSerAspPheThrSerTyrGlyProVa 588
652 .....GCATGGTTAATTGGCGGCAATACAC 676
588 lSerAsnLeuSerPheLysProAspIleThrAlaProGlyGlyAsnIle.. 604
677 ATATGAGGGGTGG.....GGAATATATGCGCTANTTAGTTGAGC 717
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
605 .....TTPSerThrGlnAsnAsnLysTyrThrAsnMetSer 617
718 GCGGATGTGCGCATGCAACGACTATGGCCCTATGCCGATTCAGAGT.. 765
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
618 Gly...ThrSerMetAlaSerPrope.....IleIalaGlySe 629
766 .....GCGGACGCGACAGCGGTTGCGCAATGTTTA 796
629 rGlnAlaLeuLeuLysGlnAlaLeuAsnAsnLysAsnAsnProPheTyrA 646
797 TTTATGCAAAACAAACAAATTAATGCTGCTCAACGAGATTACAA... 843
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
646 lAtyTyrLysGlnLeuLysGlyThrAlaLeuThrAspPheLeuLysThr 662
844 .....ACGGGCTACCTTATTCGGC..... 864
663 ValGluMetAsnThrAlaGlnProIleAsnAspIleAsnTyrAsnAsnVa 679
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
864 ..... 864
679 lIleValSerProArgArgGlnGlyAlaGlyLeuValAspValLysAla 696
865 .....AGGAAACAGCT 876
696 lAlaLeuAlaLeuGlnLysAsnProSerThrValAlaIalaGlnAsnGly 712
877 TTCAGCTGATAGC...AAAGATTGTTCTACGATGAC..... 912
713 TyrProAlaValAlaGlnLeuLysAspPheThrSerThrAspLysThrPhe 729
913 .....ATTACAGAGCGCATACACATACCGTCTGTTTACACCGCGCA 955
729 sLeuThrPheThrAsnArgThrThrHisGluLeuThrTyrGlnMetAsp 746
956 GTAACGA.....CATTTCTCTTACATCCAAACACAC 990
746 eAsnThrAspThrAsnAlaValTyrThrSerAlaThrAspProAsnSer 762
991 GGTACG.....GTAACGTTACAGAAACCAACGA 1019
763 GlyValLeuTyrAspLysLysIleAspGlyAlaIleLysAlaGlySe 779
1020 AAAGGTATCCATCCCAAGCTTAAAGTACAGACAGTCCGACTGTTTACG 1069
779 rAsnIleThrValProAlaGlyLysThrAlaGlnIleGlu...PheThr 795
1070 AATCTTGATGAATCACTGATTAAGAACCAAGTTTACGGCGAGGGGGTGT 1119
795 eUSeuLeuProLysSerPheAspGlnGlnGlnPheValGlnGlyPheLeu 811
1120 AATCAGTACCGTCAAGGTTAAACACGCTGAAACCTTTCT..... 1161
812 AsnPhe.....LysGlySerAspLysSerAlaGluAsnLeuProTy 825
1162 .....TTTATGATTTACGGCAACGCAACACTATCTTATCAACA 1201
825 lMetGlyPhePheGlyAspTyrPAsnAspGlyLysIleValAspSerLeuA 842
1202 AC..... 1203
842 snGlyIleThrTyrSerProAlaGlyLysAsnPheGlyThrValProLeu 858
1204 ..ATCAACCAAGGCGCGGGGCTTTGATTTTGAAGGTGATTTTACGCT 1250
859 LeuLysAsnLysAsnThrGlyThrGlnTyrTyr...GlyGlyMetValTh 874
1251 CTCGCCCTAAAACAAACGATGGCAAGCGCGGCGCTTCATATCACTG 1300
874 rAspAlaAspGlyAsnLysThrValAspAspGlnAlaIleAlaPheSer 891
1301 AAGCAGTACCGCTTACTGAAAGTAACGGCGGTGCGCAACGACCGCTG 1350
891 eAspLys.....AsnAlaLeuTyrAsnAspLysSer 901
1351 TCCAAA.....ATCGCAAAAGCACGCTGACGTTCAAGCCAAAGGGA 1394
902 MetLysTyrTyrLeuLeuArgAsnIleSerAsnValGlnValAsp.... 916
1395 AAACCAAGCTCGATCAGCGTGGGAGAGGTACACTCATTTTGGATCAGC 1444
917 .....IleLeuAsp.... 919
1445 AGGCAGACGATAAAGCAAAACAAAGCCTTAGTGAATGCGCTGNTC 1494
920 .....GlyGlnGlyAsnLysValThrThrLeuSerSerSer 931
1495 AGCGGACGGGTACGCTCAACTGAATCGCGATATATCAG..... 1533
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

[illegible]

2331	CGATTACA.....	CGTTATA	2343
1206	exspphrlysvaIaVgIeMeIaIePrOthrPherThraSpleuysPheh	1223	
2345	CAGTCAGCCCAACAGCCGCAACCCCAAAAGCGCAACCTTAAAGCTGGTGGCAAT	2394	
1223	snasngIygeAspGlntrIserGuaIaIaThrIleIysValaThgGlyThr	1239	
2395	GCCCAAGCAACAATTAAATCAACCCCAATTAAAC..GGCAACNCATGGGN	2441	
1240	ValSerAla.....AspIhrIystrIhValasnaValGlnsthrValaI	1254	
2442	TTGGGCAATGGCTTCATTAAATCTGAAGACACAGCCGCGCAACAAAGCGCA	2491	
1254	aalIaIeuaAspIaIaGlnIshIsIaPheSerValaAspValaIProValasn...	1269	
2492	GTTCAGCGCTTCCGCAACGCTAAAGCAACGTAAGCATTTCCGCGCATC	2541	
1270	.....TyrIeIysAspAsnThr...IleIysValaThraIaIaThrAspIys	1282	
2542	AACGCAATGTCTCCCTTACCCATAGAGCAGCATTTCCATTTTGAACAAAG	2591	
1283	AspGlyAsnThrThrIhIeIuInIshThrIle.....Thse	1295	
2592	CCGCTTTACCGGACAACATCTAGCGGCAAGCAAGANACAGCATTAACCTTA	2641	
1295	IserTyrAspProAspMet.....LeuI	1303	
2642	AACAGCAATAGGAGCGCTGAGCGAGCGAATTTAGCAATTTAAAC	2691	
1303	ysIysSerVal...ThrPheasprGlnIyValIysPheGlyThr...Asn	1317	
2692	CTTGACAAACGCCACCAATTACACTCAATTCGCGCTATCGCCAGATGCTGC	2741	
1318	LysPheAsnIaIaThr.....SerAlaIysPheTyr.....	1328	
2742	AGGCGGGCAACCGGACAGNGTGTACAGACGCGCGCGCGTTCGGGCC	2791	
1329	....ProIystrIneIyIaIaIaThrIeIhIeIyIys.....	1339	
2792	GTTCCTATTATTCGTTTACACGCCCAACTGCTGAATCCCGTTTCAAC	2841	
1340	.....ValIyshIsProThr.....	1345	
2842	AGCCTACACGGTAAAGCGCAATTTGAACNGTCAAGAACATTCGCTTAT	2891	
1346	ThrIleGlnIValaAspIyIys.....	1352	
2892	GTCGGAATCTTCGGCTACCGAAGCAACAATTTGAAGCTGGCGGAAGTT	2941	
1353	.....GlnIleProIleIysAspIle	1360	
2942	CCGAAAGNACTTACACCTTGGCGGTCAACATATCCGGCAAGCAACCGGTA	2991	
1360	euthrPheSerPheThrIeuaSpleuGlyThrIeugIyGlnIySPheProPhe	1376	
2992	AGCCTGCATCATTTAGCGTATGTGAAGGAAGCAACAAACCGCGTCTC	3041	
1377	Gly.....ValValaIaIyAspThrThrGlnAsnIystrPheG	1390	
3042	CGAAACCTTAATTTACCCCTGCA.....A	3067	
1390	ngIuIaIaIeuaSerPheIleIeuaAspIaIaValaIaIaProthrIeuaSerIeua	1407	
3068	ACGACACGTCGATCGCGCGCGGTGCTTACCAATCATTCGCAAAAGAC	3117	
1407	sPaserSerThraSplaIaProValaTyr.....ThraSnaSp	1418	
3118	GCGAGATTCGGCTGATTAATCCGGTCAAAAGACAAAGAGCTTCCGACAA	3167	
1419	ProAsnPhaGlnIle.....	1423	

33168	ACTCGGCAAGGAGGAAGCCAAAAAACAAGCGGAAAAAGCAACGGCGCAA	321
1424	.ThgLYThAla.....ThsAspAlaIagInt	1433
3218	GCCTTGACGGCGTGAATTGGCGCGCGGATCCGCGCAAAAGACAGA	3267
	:	
1433	yrLeu...SerLeuSerIleasnGlySerSerValAlaSerGlnIyrGlu	1448
	:	
3268	.....AGCGTGGCCGAAC	3281
	:	
1449	AspIleasnIleasnSerGlyLysProGlyHISmetAlaIleAspGlnPr	1465
	:	
3282	GCGCCGCGNCGAGCGGGGAAATGTCGCATTAATGACGGCGAGGAAG	3331
	:	
1465	oValLysLeuLeuGlnGlyLysasnValLeuThrVal.....	1477
	:	
3332	AGAAAAAGCGGTGACGCGCGGATTAAGACAGCGNTTGGCAACAGCGC	3381
	:	
1478	.....AlaValThrAspSerGluAspAsnThrThrThrLysAsnIle	1491
	:	
3382	GAAGCGGAAACCGCGCGGNTACACACCGCTTCCCGCGCGCGNGGC	3431
	:	
1492	ThrValIyrTyrGluPro.....	1497
	:	
3432	CGCGCGGATTTGCGCAACCGACGCCAACCGCA...CCTCAACCC	3478
	:	
1498	LysLysThrLeuAlaAlaProThrValThrProSerThrThrGluPro	1514
	:	
3479	AACCGAGCGCGACCGATNACCCGTATGCCAATAGCGGTTTGAAT...	3525
	:	
1514	IaGlnThrValThrLeuThrAlaAsnAlaAlaIaThrGlyGlnThrVal	1530
	:	
3526	GAATTTTCGCGCACGCTCAACAGCCGTTTCCGCGTACGAGCAATGCA	3575
	:	
1531	GlnTyrSerAlaAspGlyLysThrTyr.....	1540
	:	
3576	CGCGGTGTTGCCGAAGACCGCCGACGCGNTTGGCAAGCNGCATCC	3625
	:	
1541	.....GlnAspValProAlaIaGlyValThrIleThrAla	1553
	:	
3626	GGNACCAACAACTACCTTGCAAGATTTCCGCGCTACCGCAACAA	3675
	:	
1553	snGlyThrPheLysPheLysSerThrAspLeu...Tyr.GlyAsnGln	1567
	:	
3676	ACGACCTGCGCCAAATCGATATGCAAAAAAACCCTGGCAGCGGGCGGT	3725
	:	
1567	uSerProAlaValAspTyrValValIleThrAsnIleLysAla.....	1580
	:	
3726	CGGCACTCTGTTTGGCAACAACGAGCAACAAACNTTGCAGCAGCGCA	3775
	:	
1581	.....AspAspProAlaGlnLeu	1586
	:	
3776	TGCGCACTCGACAGCGCTTGCCACAGCGCGCTTTTGCGCAATACGCG	3825
	:	
1587	GlnAlaAlaLysGlnGlnLeuThrAsnLeuIleAlaSerAlaLysThrLe	1603
	:	
3826	ATCGCAGAGTTGACATCGGCATCGACAGCGCGCGGTTTACACGCG	3875
	:	
1603	uSerAlaSerGlyLysTyrAspAspAlaIleThrThrAlaLeuAlaIaAla	1620
	:	
3876	CANTC.....TNTCAGACGGCATGAGGAGCAAAATCCGCGCGCGTGC	3919
	:	
1620	IaThrGlnLysIaGlnThrAlaLeuAspGlnThrAsnAlaSerValAsp	1636
	:	
3920	TGCATTACCGCATTCAGGCACGAT.....AC	3945
	:	
1637	SerLeuThrGlyAlaAsnAspSerLeuGlnThrAlaIleasnGlnLeuAl	1653
	:	
3946	CGCGCCGCGTTTCGGGATTCGGCATCGAACCGATATCGCGCAACGCG	3995
	:	
1653	aaIaLysLeuProAlaAsp.....LysLysThrSerLeuLeuAsnG	1667
	:	
3996	CTATTTCGCCAAAAAGCGGATTACCGCTACGAAACGTCATATATCGCA	4045

1667    ::::|||||    |||    ::::|||||    ::    :::|||  
1667    IntucInserVallyAspaIalaLeuglYthrAspleuclyAynsnlrThr 16833  
4046    CCCCCGCTTGGCGTTCAACCCGNTACCNGCGGCATTAAAGCACATTT 40955  
4046    |||||    ::::|||||    |||||    ::::|||||    ::::|||||  
1684    AspProSerThrglyySthrPhePhrAla...AlaleuAspAspleuVa 16999  
4096    TCATTCAACCGGCCAACACA 4117  
4096    :    :::    ::::|||||  
1699    lAlaglnAlaGlnlAlaglYthr 1706

seq\_name: SwissProt\_40:FHAB\_BORPE

seq\_documentation\_block:

ID	FHAB_BORPE	STANDARD;	PRT;	3591 AA
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				
48				
49				
50				
51				
52				
53				
54				
55				
56				
57				
58				
59				
60				
61				
62				
63				
64				
65				
66				
67				
68				
69				
70				
71				
72				
73				
74				
75				
76				
77				
78				
79				
80				
81				
82				
83				
84				
85				
86				
87				
88				
89				
90				
91				
92				
93				
94				
95				
96				
97				
98				
99				
100				

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-FEB-1996 (Rel. 33, Last annotation update)

GN FHAB.

OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;

NCBI\_TaxID=520;

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE

RA Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;

RT haemagglutinin: a protein processed from an unusually large

RL Mol. Microbiol. 4:787-800(1990).

RP SEQUENCE OF 1-3261 FROM N.A.

RA Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;

RT sequence and crucial role in adherence.";

CC -!- FUNCTION: EVIDENCE FOR A ROLE IN HOST-CELL BINDING AND

CC -1- SUBCELLULAR LOCATION: SURFACE

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC the European Bioinformatics Institute. There are no restrictions on its

CC modified and this statement is not removed. Usage by and for commercial

or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL; M60351; AAA22974.1; -.

EMBL; M60351; AAA22976.1; ALT\_INIT.

SEQUENCE	AA	MW	EF7418B30D6E5138	CRC64
3591	AA	367420		

```

--format=tsv:
Quality: 180.50
Length: 1427

```

macro:	0.000	gaps:	7
Percent Similarity:	41.416	Percent Identity:	19.131

alignment\_block:

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

..... 309 1/1 00: THREE-DONE 2 10000: 1 00: 3091

202 CCCCCCIIIIIINIAACAAACAAAAAGGGAG..... 23/

2200 KATOLNOLY DEUOTIUVARJEBELYNALYNASIIAAGLIIVAIHIAASPAL 2300

1



2738 uGlyAlaLysValGlyAla.....GlyLysT 2747  
1859 CAATTATTAAGCGCAATCATCGCAAAACAAAGCAAACTGTTTTCAGC 1908  
2747 YrGluAlaGlyPheSerLeuGlySerGlyLeuGlyAlaHisAla 2763  
1909 GGCAGACCGCAGCAGCGCCTTACATCATTTAGGAAGCGGTGTGCTAAA 1958  
2764 GlyArg..... 2765  
1959 AATGGAAGTATCCACAGAGAAATCGT..... 1989  
2766 .....GlyMetThrAlaGlyAlaGlyValGlyTyrArgAla 2780  
1990 .....TGGACAACAGACTGATGATCAGGCACTTAAAGCGGAATTTTC 2034  
2780 eRhISgluInSerSerGluThrGluSerGlyTyrArgAlaHisAla 2796  
2035 CATATTCAGGGCGGCGAG..... 2052  
2797 AsnPheGlyGlyGlySerValGlyAlaGlyLysValLeuAspIleGly 2813  
2053 .GCGGTGATTCCCGCAAT..... 2070  
2813 yAlaAspIleAsnArgAsnArgTyrGlyGlyAlaAlaLysGlyLysAla 2830  
2071 .....GTTGCCAAAGTGAAGGCGATTTGN 2094  
2830 LyThrGluGluAlaLeuArgMetArgAlaLysLysValGlySerThrLys 2846  
2095 CATTGAGCAATACAGCCCAAGAGATTGTTGTCGACCGCATCAAG 2144  
2847 TyrValSerGluGlnThrSerGlnSerSerGlyTyrPheValGlyValAl 2863  
2145 CCATACATCTGTACGTTGCGACTG.....ACNGTCTGA 2182  
2863 aSerThrAlaSerAlaArgSerSerLeuThrAlaAlaThrArgLeu 2880  
2183 CAATTTGTGCAANAAACATTCACGACGATAA..... 2217  
2880 LysPheSerValAlaGlnAsnValGlnAspArgGluIleArgGly 2896  
2218 GTGATTGCTTATTGACTAGACAGACNTNAGCGGCAANTGNTNMC 2267  
2897 LeuMetAlaAlaGlnValAlaAlaGlnAlaThrGlnLeuValThrAlaAs 2913  
2268 CNAATNAGTNTNTNAAANCTCNGGCGNTGCAACNCACTNANAGCATC 2317  
2913 P.....ThrAlaAlaValAlaLeuSerAlaGly 2923  
2318 TTAGTGCATATGCGATACAGTTATACAGCCACAAAGCCACCAA 2367  
2923 LeSerAlaAspPheAspSerSerHis.....SerArgSerThrSerGln 2937  
2368 AAC.....GCCAACCTTACG..... 2382  
2938 AsnThrGlnTyrLeuGlyGlyLysLeuSerIleGluAlaThrGlnGlyAs 2954  
2383 .....CTCGTGGG..... 2391  
2954 PAlaThrLeuValGlyAlaLysPheGlyGlyAspGlnValSerLeu 2971  
2392 .....AATGCCAAGCAACATTTAAACAA 2415  
2971 yAlaAlaLysSerValAsnLeuMetAlaAlaGlnSerThrPheGlnSer 2987  
2416 GCCCATTTAAAGCGCAACNCATCGGNTTCGGCAATGCTTCATTATCT 2465  
2988 TyrSerGlnSerHisAsnPheHisAlaSerAlaAspAla.....AsnLe 3002  
2466 AAGCAACAACCGCGCACAAACAGCAGTCAGCTTCGACAAACGCTA 2515  
3002 uGlyAlaAsnAlaValGln...GlyAlaValGlyLeuGlyLeuThrAlaG 3018  
2516 AGCGAAGCTAAGCCATTCGCGACTCAAC..... 2544  
3018 LyMetGlyThrSerHisGlnIleThrAsnGlyThrGlyLysThrTyrAla 3034  
2545 .....GCAATGCTTCCTTA.....GCGCATTAAGCAGT 2573  
3035 GlyThrSerValAspAlaAlaAsnValSerIleAspAlaGlyLysAspLe 3051  
2574 ATTCCATTTTGAAGAACGCGCTTACGCGACACTACGCGGACAGAA 2623  
3051 u.....AsnLeuSerGlySerArgV 3058  
2624 ANACCATTAACCTTA..... 2640  
3058 aLArgGlyLysHisValValLeuAspValGlyLysPheAsnAlaThr 3074  
2641 .AAGACAGCGA.....TGAGCGT 2660  
3075 SerLysGlnAspGluArgAsnTyrAsnSerSerGlyGlyTyrPheAspAl 3091  
2661 GCCGTCAGCAGCGAATTAGCAATTTAAACCTT..... 2694  
3091 aSerAlaGlyValAlaIleGlnAsnArgThrLeuValAlaProValGly 3108  
2695 .....GACACGCGCACATTAACACTC 2715  
3108 eRhAlaGlyPheAsnPheAsnThrGlnHisAspAsnSerArgLeuThrAsn 3124  
2716 AATTCCGCTATTCGCCAGATGCTGAGCGGCAACCGGAGNGTGTG 2765  
3125 AspGlyAlaAlaGlyValAlaAlaSerAspGlyLeuThrGlyHisVal 3140  
2766 AGACAGCGCGCGCGCGTTCGCGCTTATATCCGTTACACCGC 2815  
3141 .....LysGlyAspAlaAsnLeuThrGly 3149  
2816 CAATTCGCTAGAA...TCCCGTTTCAACAGCTGACGTAACGCAAA 2862  
3149 LeThrIleAlaAspLeuSerGlyLysGlyAsnLeuLysValAspGlyAla 3165  
2863 TTGACNGTCAAGAACATTCGCTTATGTCGGAACCTTCGCGCTACCG 2912  
3166 ValAsnAlaGln.....AsnLeuLysAspTyr 3175  
2913 AAGCGAACAATTGAACCTGCGGAAGTTCGGAAGNACTTACACTTG 2962  
3175 g...AspLys.....AspGlySerGlyGlyLeuAsnValG 3187  
2963 CGGTCAACAATACCGGCAACGAAACCGTAAAGCTGATCAATTGACGTA 3012  
3187 LyLeSerSerThrThrLeuAlaProThrValGlyAlaAlaPheGlyArg 3203  
3013 GTGGAAAGGGAAGACAACAAACCGCTGTCGGAACCTTAATTACCT 3062  
3204 ValAlaGlyGlnAspTyrGlnAlaGlnAlaArgAlaThrIleAspValG 3220  
3063 GCAAAAGCAACAGCTGATGCGCGCGCT.....GAGCGT 3097  
3220 LysIleThrLysAsp.....ProAlaArgLeuGlnValGlyGlyVal 3234  
3098 ACCAATCATTCGCAAGAGCGGAGTTCGCTGATATCCGTCANA 3147  
3235 LysGlyThrLeuAsnGlnAspAlaAlaGlnAlaThrValGln...Ar 3250  
3148 GAACAAG.....AGCTTTCCGACAACCTGGCAA 3176  
3250 gAsnLysHisThrPheAlaGlyGlySerGlnPheSerValAlaGlyLys 3267  
3177 GGCAGAAGCCAAACAAACAGCGGAAAGAACAAAGCGCAACGCTTACG 3226  
3267 eRhLeuLysLysAsnGln..... 3273





```
762 AGTGGCGGACGACAGCGGTTGCCAATGTTTATTATGCAAAAACA. 810
    :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
153 yglValaIleTyrSerGlnGlyIleuLeuPheThr...AspLeuThrG 169
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
811 .....ACAAATTAATGGCTGCTCAACGAGATTTTACAACC 846
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
169 IyLeuThrIleGlnGlyAsnLeuSerGlnLeuSerGly..... 181
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
847 GGTACCCCTTATCCGCGAGGAGAAAAGGTTTCCAGCTGATACGCAAGA 896
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
182 GlyIleIlePheGlyGlySerThrIleSerPheSerGlyIleAsnGlnAl 198
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
897 TTGGTTCTACGATGACATTTACAGAGCGCATACATACCGCTGTTTGTG 946
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
198 aThrPhe.....SerSerAsnThrAlaGluVal 208
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
947 AACGGCCGACGTAAAGGACATTTTCTTTCATCCAAACAAACAGGTACG 996
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
208 aProGluGlu.....ThrThrProAsnProAsnPro 218
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
997 GGTACGTTACAGAAACCAAGCAAGAAAGTTTCCAAATCCAAAGTTAAAGT 1046
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
219 GlyThrGlnThrThrThrSerGln.....ProSerProThrSerLysVa 233
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1047 ACAGACAGTCCGACTGTTTGACGATCTTGAATGAATACTGATTAAGAAC 1096
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
233 IGIInSerLeuPheThrTyrSerSerSerThrGlnAlaAsn..... 246
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1097 CAGTTTACGGGCGGCGGTCTTAAATCAGTACCTCCAGAGTTAAACAAAC 1146
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
247 .....GlyAsnGlyAlaAspSerGlnThrProSerHisLysPro 259
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1147 GGTAACAAACCTTTCTTATGATTTACGCAACGCGCAAACTCAATCTATC 1196
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
260 GlySerGlyGlyAlaIle.....TyrAlaThrGlyAspLeuThrIleSe 274
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1197 A.....AACAACTCAACCAAGCGCGG 1219
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
274 rAspSerGlnGluIleValPheSerValAsnLysAlaSerLysAspLys 291
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1220 GCGGTTGTAT.....TTTGAAGGTGATTTTACGCTC 1251
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
291 IyAlaIlePheAlaGluLysAsnValSerPheGluAsnIleThrThrLeu 307
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1252 TCGGCTGAAAAACAAGCAAGCGTGACAA...GGCGGCGGCGTTGCATATCAG 1298
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
308 LysValAlaIleAsnAsnGlyAlaGluGlyGlyGlyIleTyrAlaSe 324
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1299 TGAAGACAGTACCGTCTTGAAGAGTAAAGCGCGTGCAAAAGCGCGCC 1348
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
324 rGlyAspLeuSerIleGlnSerLysGlnSerLeuPheAsnSerAsnTr 341
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1349 TGTCCAAATTCGCAAGGCAAGCGTGCATTCAGCC..... 1386
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
341 hrSerLysGlnGlyGlyAlaLeuThrIleGlnGlyAsnValaAspPhe 357
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1387 .....AAAGGGAAACCAAGCGTCGATCAG 1412
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
358 LysAspLeuGluGluIleArgIleLysTyrAsnLysSerGlyThrPheG 374
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1413 CGTGGCGGCGTACAGTCACTTTTGCATCAGAGCGAGACGATAAAGGCA 1462
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
374 urThrLysLysValThrLeuSerLeuProGluAlaGlnThr..... 387
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1463 AAAAACAAGCTTTAGTGAATGGCTGTGTCAGCGGCGAGGGGTACGGTG 1512
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
388 ...AsnLysSerSerVal...ThrAlaAlaSerGlnSerGly..... 399
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1513 CAACGTGAATCCGATATCAGTCAACCCCGACAAACTATATTTCGGCTT 1562
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
400 .....ProAsnThrThrProThrProThrProThrProThrAlaAla 412
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1563 TCGCGGCGGACGTTTGATTTAAACGGGCACTTGCTTTCACCCGTA 1612
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

413 .....LysGlyGlyGlyLeu..... 417
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1613 TTCAAATATCCGATGAAGGCGGATGATTCGATTCATTAATGACACACA 1662
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
418 .....TyrThrGlnLysAsnLeuSerIle.....SerAsnIle 428
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1663 ACATCCACCGTTACCTTTACAGGAAATGAAGTATTTACACAAACCGAGTG 1712
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
429 ThrGlyIleIleGluIleThrAsnAsnLysAlaThrAspValGlyGly 445
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1713 T.....ACAAATATCAATACGTTAAAT 1735
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
445 yAlaTyrValLysGlyThrLeuThrCysLysAspSerHisAlaGluGlnP 462
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1736 ACAGCAAGAATAATGGCTACAAACGGTGGTTGGCGAGAAAGATACAGCC 1785
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
462 he.....GlnLysAsnSerSer 467
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1786 AAAAGCAAGCGGCGGCTCAACCTTGTTTACAGCCCGCGGAGAGACCG 1835
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
468 GluLysLysGly..... 471
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1836 CACCCNGCTGCTTTCGCGGGAACAATTTAAACGGCAACATCAGCAAA 1885
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
472 .....GlyGlyLeuTyrThrGlnAspThrIleThrLeus 483
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1886 CAAC....GCCAACTGTTTTCAGCGGCGAGACCGACCGACCGCC 1929
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
483 erAsnLeuThrGlyLysThrLeuPheGln..... 492
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1930 TCAATATCATTTAGGAAGCGGTGTCAAATAATGGAAGGATCCCAAGG 1979
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
493 .....GlnAsnThrAlaLysGluGluGly.....GlyG 502
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1980 AGAAATCGTGTGGAGCAACGACTGATCCNACCGCATTTAAACGGGAAA 2029
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
502 yLeuTyrIleGlnGlyAspAspLysThrLeuThrIleThrGlyLeuAsp 519
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2030 ATTTCATATTT.....CAGGGCGGCGAGCGGTG 2058
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
519 erPheCysLeuIleAspAsnThrSerAlaThrHisGlyGlyAlaTyr 535
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2059 ATTTCGCGAATGTGCCAAATGCGAAGCGCATTCGATTCAGCAATCA 2108
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
536 ValThrLysGluIleSerGlnThr.....TyrThrSerAs 547
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2109 CCGCCAGACGATTTTGTGTGTCACCGCATCAAGCATACATCTGTA 2158
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
547 rValGluGluPheProGlyIleThrProValHisGlyGluThrIleIle 564
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2159 CACGTTGAGCTGACGNGCTGCAAAAT.....TGTGTGCAANAA 2199
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
564 erGlyAsnLysAlaThrGlyGlySerGlyGlyValaCysThrLysHis 580
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2200 AHCATTTACGAGATTAAGTGAATGCTTCATTCAGTAAG.....AC 2240
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
581 LeuValLeuSerAsnLeuGlnThrIleSerIleSerGluAsnPheAlaSe 597
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2241 NCACNTNAGCGGCANTGTNAGNCTNNCANTNAGCNTNNTTNAANCNTN 2290
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
597 rGluAsnGlyGly..... 601
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2291 CNGGCGTNCNCACTNAANGCAATCTTAGTGAATAAGCGGATACAGCT 2340
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
602 ...GlyAlaCysThrCysProAspAsnPheProAla.....ProThrAla 615
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2341 TATACGTACGACACACGCAAC.....CAAAACGG 2372
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
616 SerThrProSerThrAsnGlnThrAlaAlaProLysAspAspLysAspPh 632
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2373 CAACCTAGCCCTGTGGCAATGCCCCAAGCAACATTTATACGCGACAT 2422
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
```

```
632 eleuileasptyrvalvalserthrthrlleasplysasnlvalarhl 649
2423 TAAACGGCAACNCATCGGNTTCGGGCAATGCT..... 2454
649 ysllysglyalaglyvaltyralalyslysalyserargileasp 665
2445 TCATTATATCAACCAACAGCCGACAA.....AACGGAGCTGAC 2498
666 gluleuasnlleserapasnalaalaglnlthrlyglylyglypnecy 682
2499 GCTTTCGACAAAGCTAAGGCAAAAGTAAGCATTCGCACTCAACGGCA 2548
682 scythrthrgluserleugluleuasn..... 690
2549 ATGTCCTCCCTCGCATTAAGGCAATTCATTTTGAAGACGGCTTT 2598
691 .....Thrllealaserleuaserval..... 697
2599 ACCGGACAACCTACGCGCAGCAGCAAGANACAGATTACCTTAAAGACAG 2648
698 Thrlysasnlleualaglylyslyglylyglyleuhtsalys..... 712
2649 CGAATGGACGCTGCCGTCAAGCAGCAATTAAGCAATTAAACCTTGACA 2698
713 .....Thrlleuasnllesera..... 718
2699 AC.....GCCACCATTAACCTCAATTCGGCTATCGGCAGAT 2736
718 snleuylserserlyleuserpheserasnthrtralaasnseraser 734
2737 GCTCAGCGCGCCAAACCGCAGNGTGTACAGACGGCCGC...CCGCC 2782
735 Thrlygvalalathrthrthralathrthrserserprothrvalserse 751
2783 GTTCGCGCGCTCCCTATATCCGTTCACCGCCCACTCGGTGATCC 2832
751 rphelueupro.....Argalathrthralaglyserseproal 764
2833 CGTTTCAACAGCGTGCAGGTAAGGCAAAATTAACNGTCAAGACAT 2882
764 roalagln.....Thrthr 768
2883 CCGCTTATATGCGAACTCTTCGGCTACCGAAGCGCAAAATTAAGCTGG 2932
769 Prothrlyrlyalaglyval..... 774
2933 CGGAAGTTCGGAAGNACTTACACCTTGGCGGTCAACAATACGGGCAAC 2982
775 .....Valglylyalaliletyr.gly... 781
2983 GAACCCGTAAGC.....CTCGATCAATGACGGTAGT 3014
782 Gluthrvalaserpheserlyscysserglyleucysglinphethrtrguals 798
3015 GGAAGGGAAGAAACAACAAACCGCTGTCCGAAACCTTAATTTCACCTGTC 3064
798 nseralalileaspsanthrproserseproserleuasn..... 811
3065 AAAACGACACGTCGATCCGCGCGCTGACCACTCATTCGGCAAA 3114
812 .....Valglnglylyalalaliletyr.alaly 820
3115 GACGCGAGTTCGGCTCGCATATCCGGTCAAGAACAGAGC..... 3157
820 sthrserleuser.....Ileglualaglnasproserthrsertr 834
3158 .....TTTCGCAACAACCTCGGCAAGGCAAGACCAAAAAACAGCGGAGAA 3202
834 yvalalpheaserlysasnserval..... 841
3203 AAGACACAGCGCAAAAGCTTGACGCGCTGATTCGGCGGCGGCGGATGCC 3252
842 .....Serthrlyglylyalaglnthr..... 848
3253 GCGGAAAGACAGAAAGCTTCGCAACCGCGCGCGCAGCGGCGG 3301
849 .....Thrglylnlealaglyg 855
3302 .....AAATATCGGCAATTATGACAGCGGAGAGAGAGA 3334
855 lyalaliletyrserproservalthrleuasncys..... 866
3335 AAAAAGGTTGACGCGGATTAAGACAGCGCTTTGGCCAAACAGCGGCA 3384
867 .....Glnthrvalalpheaserlysasnservalase 876
3385 G.....CGAAACCGCGCGGNTTACACCGCTTCGCCCGCGC 3422
876 rmetalathrthrtrasnproproserglythrserprolysasphrileg 893
3423 CCGCNCGCGCGCGCGGATTTCCGCAACCGCAGCGCCCAACCGCAACTC 3472
893 lyglyalalilealaglythrthrthrlleaserlyserlyserthrsphs 909
3473 AACCCCAACCGCAGCGCGACCTGATNAGCGTTATGCCAATAGCGGTTG 3522
910 Sergluasn..... 912
3523 AGTGAATTTTCGCGCAGCTCAACAGCGTTTTCGCCGTACAGGACGAAT 3572
912 ..... 912
3573 GGAACGCGTGTTCGCCAGAACCCCGCAGCGCTTTGGACACAGCNCA 3622
913 .....Thrlalaspleu.....Glyalalalileglythrleuser 924
3623 TCCGNAACCAACAACACTACGCTGCGCAAGATTTCGCGCGCTTAC... 3667
925 glylyserseaserasnleuthrlystlethrleuserasnlyse 941
3668 .....GCCAACAA.....CCGACCTGCGCA 3689
941 rphethrthpheglulysasnlyalalasnlysarlyglyaliletyralap 958
3690 AATCGGTATGCAGAAA.....ACCTCG 3712
958 roservalaserlilelysglyasnasnllethrphesnglnasnthrser 974
3713 GCGCGCGCGCGCTGCGCATCTGTTTTCGCAACCGCAGCGCAAAACANC 3762
975 Thrhtlaspeylseralaliletyrphs..... 983
3763 TTTCGACAGCGGATCGCAACTCGGCACGCTTCCCGCAGCGCGCTTT 3812
984 .....ThrlysasplalathrthrllegluserleuglyserValleuphet 998
3813 CCGGCATATACGGCATCGCAGGTTCGACATCGGCATACACAGCGCGCG 3862
998 hrclyasnasnvalThrtralaglnalalasersethr..... 1010
3863 GTTTTACAGCGCGCANTTNTCAGACGCGCATCGGAGGCAAAATCCGCGC 3912
1011 .....Alathrlyglylgnthrthrtrasnthrthraspyrlyal 1024
3913 CCGCTGCTGCATTAAGGCAATTCAGGACAGATACCGCGCGGTTTCGCGC 3962
1024 aala.....IlephaglyA 1029
3963 ATTTCGCATTCGAACGCTTACA...TCGGCGCAACCGCTTTCGTCCAAA 4009
1029 sptrogllythrtrthrtraspthrthrlleuasnllealaserse 1045
4010 AAGCGATTACGCTACGGAAC.....GTCAATTCGCCAC... 4047
1046 glyasnllethr..Pheserasnasnsersglnasnthralathrtrasn 1061
```







RC STRAIN-IFO 3046;  
 RX MEDLINE=86223815; PubMed=3011754;  
 RA Yanagida N., Nozumi T., Beppu T.;  
 RT Specific excretion of *Serratia marcescens* protease through the outer  
 RT membrane of *Escherichia coli*.  
 RL J. Bacteriol. 166:937-944(1986).  
 RN [2]  
 RP PARTIAL SEQUENCE, AND PROCESSING.  
 RX MEDLINE=92348352; PubMed=1639760;  
 RA Shikata S., Shimada K., Kataoka H., Hornouchi S., Beppu T.;  
 RT Detection of large COOH-terminal domains processed from the  
 RT precursor of *Serratia marcescens* serine protease in the outer  
 RT membrane of *Escherichia coli*.  
 RL J. Biochem. 111:627-632(1992).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE  
 CC SUBTILASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M13469; AAA26572.1;  
 DR PIR: A28840; A29840.  
 DR HSSP: Q99405; IMP.  
 DR MEROPS: S08.094;  
 DR InterPro: IPR000209; Peptidase\_S8.  
 DR Pfam: PF00082; Peptidase\_S8; 1.  
 DR PRINTS: PR00723; SUBTILASIN.  
 DR PROSITE: PS00136; SUBTILASE\_ASPI; FALSE\_NEG.  
 DR PROSITE: PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE: PS00138; SUBTILASE\_SER; 1.  
 DR Hydrolyase: Serine protease; Zymogen; signal.  
 FT SIGNAL 1 27  
 FT CHAIN 28 645 EXTRACELLULAR SERINE PROTEASE.  
 FT PROPEP 646 1045  
 FT ACT\_SITE 76 76 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 112 112 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 341 341 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 SQ SEQUENCE 1045 AA; 112345 MM; 4924EAS0E4F179C CRC64;  
 -----  
 alignment\_scores:  
 Quality: 177.50 Length: 1190  
 Ratio: 0.362 Gaps: 56  
 Percent Similarity: 41.176 Percent Identity: 19.160  
 -----  
 alignment\_block:  
 US-09-303-518D-651 x PRTS\_SERMA ..  
 -----  
 Align seg 1/1 to: PRTS\_SERMA from: 1 to: 1045  
 -----  
 1135 AGGTAACACACGGT.....GAAACCT 1157  
 91 ArgLeuAspAsnGlySerTyrAsnPheserTyrAspLysGlnAspAsnMe 107  
 1158 TTCCTTTATGCGATTACGGCAACGCAACTCATCTTATCAACAACATCA 1207  
 107 tSerPheGlyAspHisGlyThrHisValAlaGlyIleHisAlaIleLysA 124  
 1208 ACCAAGCGCGCGGC.....GTTTGTATTGTAAGGATTTTACGGTC 1251  
 124 TGASPGIAlaGlyMetHisGlyValAlaPheAspAlaAspIleIleGly 140  
 1252 TCGCCTGAAGC..... 1263  
 141 ThrIleuAsnAspTyrGlyAsnArgAsnGlyArgGluGluLeuIleG 157  
 1264 .....AACGAACGCTGGCAAGCGCGCGCTTCATA 1294

157 nSerAlaAlaArgValIleAsnAsnSerTrp...GlyIleAlaProAsp 173  
 1295 TCAGTCAGACACAGT.....ACCGTTACTGGAAAGTAAAGGC..... 1332  
 173 LArgArgAspAlaLysGlyAspIleIleTyrLeuProAsnGlyArgPro 189  
 1333 .....GTGGCAACGACCGCGCTGTGTCAAATCGGCAAG 1367  
 190 AspTyrValAlaPheValLysSerGluValIleAlaGluMetMetArgSe 206  
 1368 CACGCTGACGTTCAAGCCAAAGGGAACCAAGCCTCGATCAGCTGG 1417  
 206 LysSerSerValGlu...TyrGlySerGluGlnProValProThrGlyG 222  
 1418 GCGACGCTACATCTTTGGATCAGCGCAGCAGATAAAGCAAAAA 1467  
 222 LHisSerAlaMetSerThrLeuArgAlaAlaArgHisGlyLysLeu 238  
 1468 CAAGCCTTAGTGAATCGCTTGNTCAGCGCAGGAGGATCGTGCAACT 1517  
 239 LLevalPheSer.....AlaGly..... 244  
 1518 GAATGCGCAATATCAGTTCAACCCGAC.....AACTCTATTTCG 1558  
 245 AsnTyrAsnAsnTyrAsnIleProGluValGlnLysSerLeuProTyrA 261  
 1559 GCTTCGCGCGCGACGTTTGATTAACGGGCAATCGCTTCGTTCCAC 1608  
 261 LAspProAsp.....ValLeuAsnAsnTyrLeuIle..... 271  
 1609 CGATTCAAAAT...ACCGATGAGGGCGGATGATGNCNATCAATATGC 1655  
 272 ...ValThrAsnLeuSerAspGlnAsnGlnLeuSerValSerSerThse 287  
 1656 C.....ACAACAATCCACCGTTACATTACAGGAATGA.... 1692  
 287 rcysGlyGlnThrAlaSerTyrCysValSerAlaProGlySerAspIleT 304  
 1693 .....AGATTACCAACGAGGAGTGAAGTATCAATATGA 1728  
 304 ySerThrValGlyArgLeuGlnSerAsnThrGlyValAlaValAsnArg 320  
 1729 CTATATACAGCAA...GAAATGGCTTACAAACGTTGTTGGCGAGAA 1775  
 321 GluAlaTyrAsnLysGlyLysLeuSerLeuAsnProGlyTyrGlyAsnLys 337  
 1776 AGATACGACCAA..... 1788  
 337 sSerGlyThrSerMetAlaAlaProHisValThrGlyValAlaValAlL 354  
 1789 .....ACGAACGGGCGGTCAACCTGTTTAC 1815  
 354 euMetGlnArgPheProTyrMetSerAlaAspGlnIleSerAlaValIle 370  
 1816 CACCCCGCGCAGACAGACCGCACCCNGCTGTTCCGGCGGAACAATT 1865  
 371 LysThrThrAlaThrAspLeuGly.....ValAlaGlyIleAspAsnLe 385  
 1866 AAACGCG.....AACATCAGCAACA...AACGCG...AAC 1897  
 385 uphGlyTyrPheGlyArgValAsnLeuArgAspAlaIleAsnGlyProLysM 402  
 1898 TGTTTTACGCGGACAGACGACAGCGACGCTTACATATTTAGGAAC 1947  
 402 etHeIleThrLysGluAspIleProGlnGluTyr..... 413  
 1948 GGGTGTCAAAATGGAAGGTATCCCAAGAGAAATGCTGTGGACAA 1997  
 413 ..... 413  
 1998 CGACTGATCNCACGACGTTTAAAGCGAAATTTCCATTATTCAGGGCG 2047  
 ..... 2047

```
414 ....TyrValProGlySerTyrSerGluLysGlnPhe..... 424
2048 GGCAGGCGGTGATTCGCCCAATGTTGCCAAGTGGAGGCGATTGNCAT 2097
425 .....ValValAsnIleProGly..... 430
2098 TTGAGCAATCAGCCCCAAGCAAGTTTGGTGGCCAGCCGATCAAGGCA 2147
431 LeuGlyAsnIleValGlnProGlyThrProValGluArgGlyThrSe 447
2148 TACAACTGTGACAGCTTGGAGCTGGACGNGCTGCACAAATTGTGCGAN 2197
447 rSerGlyCysSerPheAspSerTyrPser..... 456
2198 AAACATTACGCGAGATTAAGTGGATTGCTCATTTAGTAAAGCAGCMTN 2247
457 .....AsnAspIleSerGlyHisGlyGlyLeuThrLysThrGly... 469
2248 AGCGGCATGTAGCNCNTNCCNATNAGCMTNNTNAAANCCTCNGGCGCN 2297
470 .....AlaGlyTh 472
2298 TGCNCACTNAAAGCAATCTTAGTGAACATGCGATACAGTTATACAG 2347
472 rLeuAlaLeuLeuGlyAsnAsnThrTyrArgGlyAspThr..... 485
2348 TCAGCCACAAGCCCAAGCAACGCAACCTTAGCCTGGGCAATGCC 2397
485 ..... 485
2398 CAAGCAACATTTAATCAAGCCATTAAGGCAACNCATCGGNTTGGG 2447
485 ..... 485
2448 CAATGCTCATTTAATTAAGCAACAGCCGCAAAAGGAGTGTGA 2497
486 .....TyrAlaLysGlnIleValLeuAla 493
2498 CGCTTCCGACACGCTAAGGCAACGTAAGCCATTCCGACATCAACGCG 2547
493 IalIeAspGlySerValAlaSerAsnVal..... 502
2548 AATGCTCCCTAAGCGATTAAGGAGTATTTGAAACAGCCGCTT 2597
503 .....TyrIleGluAsnSer..... 507
2598 TACGGAACACTACGCGGACAGCAAGACATTAACATTAAGAACA 2647
508 ....GlyThrLeuSerGlyGluGlyThrValGlyAlaPheArgAlaAla 523
2648 GCGAATGAGCCTGCCCTGACGAGCAAGCAATTAAGCAATTTAAAC...CTT 2694
523 rSerGlySerValAlaIarProGlyIleGlyThrLeuIleValLeu 539
2695 GACAAGCGCCACATTAACATTAATTCGCGCTATGCGGACGATGCTGAGG 2744
540 HisAspAlaIlePheAspArgGlySerGlnTyrAsnValGluValAlaAs 556
2745 CGCGCAACCGCGCAGNGTGTACAGACAGCCGCGCGCTTCGCGCGCTT 2794
556 P..... 556
2795 CCCTATTATTCGTTACACCGCCAACTCGGTAGAACCCGTTTCAGACG 2844
556 ..... 556
2845 CTGACGGTAACGGCAATTTGAACGTCAGAGACATTCGCTTTATGTC 2894
557 .....AsnGly..... 558
2895 GGAACCTTTCGCTACGGAAGGCAAAATTTGAACCTGGCGAAAGTTCGG 2944
559 .....ArgSerAspLysIleAlaIleAlaIleArg..... 567
2945 AAGNACTTACACCTTGCGGTCAACAATACCGGCAACGACCCGTTAAC 2994
568 .....AlaPheLeuAsnGlnIleGlySerValAsnValSer 578
2995 CTGATCAATTTGACGATAGTGGAAGGGAAGACAAACAAACCGGTGCCA 3044
579 LeuGluArg.....SerG 583
3045 AAACCTTAATTTACCTGCAAAACGACACGTCGATGCCGGCGGTGGC 3094
583 AsnLeu.....LeuSerGlnAsnGlnAlaGlnSerLeuLeuGlyAsnI 598
3095 GTTACCAACATCATCCGCAA.....GACGCGAGTTCGCGCTG 3132
598 ystYrThrIleLeuThrThrAspGlyValThrGlyArgPheGluAsn 614
3133 CATTAATCCG.....GTCAAGAACACAGAGCTTTCGA 3164
615 AlaAsnProSerTyrProPheValLysValAlaLeuAspTyrArgGlyAs 631
3165 CAACCTGCGCAAGGCAAGCAACCAAAACAGCGGGAACACACGCGC 3214
631 AspValGlyLeuGlyIleThrArgThrAspAlaSerPheAspSerLeu 648
3215 AAAGCCTTGACGGGCTGATTGGCGCGCGCGCGATGCCGCGCAAAAGCA 3264
648 IasThrThrGluAsnGlnLysAlaValAlaArg.....AlaValGluThr 662
3265 GAAAGCCTTGCGCAACCGCGCGCGCGCGCGGGAATTCGCGCAT 3314
663 LeuAsnAlaThrGlnProValThrGlnThrAlaLysArgSerValAlaI 679
3315 TATGCAAGCGGAGAGAGAAAAACGGGTGCAGCGCGATTAAGACAGCG 3364
679 e...ProAlaIleGlnIleValAsnLeuLeuGlnSerAspGlyGlyIle 695
3365 CNTTGGCAACAGCGGCAAGCGGAAAC.....CGCGCGGTACG 3405
695 IeGlnAlaValAsnGlnIleValAsnIleValAlaGlnIleProIleTyr 711
3406 ACCGCTTCCCGCGCGCGCGCGCGCGCGGATTTGCCGCAACGCGA 3455
712 GluSerPheLeuGlyPheThrSerAlaIleArgGlu.....LeuGlnI 725
3456 GCCCAACCGCAACCTCAACCCCAACCGCGCGCGCGATTAACCGCTT 3505
725 AlaThrArgGlnLeuSerGlyGlnIleHisAlaAspMetAlaSerAlaG 742
3506 ATGCAATACGCGTTGAGTGAATTTCCGCCACGCTCAACAGCCGTTTC 3555
742 IlnIleAsn.....GluSerAlaGlyTyr 748
3556 GCCGTACAGAGCAATTTGACGCGGTG...TTTCCGGAACCGCGCGC... 3600
749 LeuArgAspThrAlaThrGlnArgLeuArgGlnAlaGlnIleYArgArgTh 765
3601 .....AAGCNGTTTGGACAACGNCATCC 3625
765 rAlaThrAspIleLysAlaAspAspAsnGlyAlaThrAlaLysLeuLeuG 782
3626 GGNACACCAACACACTACGTTGCAAGAT...TTCCGCGCTACCGCCAA 3672
782 LysThrProGlyHisAlaSerGlyAsnAspAsnAlaThrGlnIleYArgGlnThr 798
3673 CAACCGACCTGCGCCAAATTCGTTATGCAAAAAACCTC...GCGACGCG 3719
799 SerThrTyrGlyValLeuLeuGlnIleLeuAspSerGlnLeuPheGlyAsp 815
3720 GCGCGTGGGCAATCCGTTTTCGACACACGCGCAACCGGAAACATTCGACG 3769
815 yArgLeuGlnYleuMetThrGlyTyrThrArgThr.....SerLeuAspG 830
```

```
3770 ACGGCATCGGCACTCGGACGGGCTTGCCACGCGGCCGCTTTTCGGGCA 3819
      ||| .....|||::: ::::: |||
830 1yGlyTyrGlnSerAspAlaHis...SerAspSntYrHisLeuGlyLeu 845
      |||
3820 TACGGCATCGGACAGTTTCGACATCGGATCGACGCGGCGGTTTTCAG 3869
      |||||
846 TyrGlnAspLysArgPhe..... 851
      |||
3870 CAGCGGCANTCTNTCAGACGCGCATCGAGGCAAAATC..... 3906
      |||:::|||:::
852 ....GlyAlaLeuAlaLeuAlaGlnGlyGlyThrTyrThrTrpHisArgI 867
      |||
3907 ....CGCGCGCGGCTGCTGCATTACGGCATTCAG..... 3936
      ||| :::::||||| |||
867 leAspThrSerArgSerValAsnTyrGlnAlaGlnSerAspArgGlnLys 883
      |||
3937 GCACGATACCGCGCGGCTTTCGGGAGTTCGGCATCGAACCGGTACATCGG 3986
      |||:::||||| ||| ::::: |||||
884 AlaLysTyrAsnAlaArgThrGlnLeuPheIleGlu..... 896
      |||
3987 CGCAACCGCGCTATTGCTCCAAAAGCGGATACCGCTACGAAACGTCA 4036
      ::::: |||::: |||:::
897 .....SerGlyTyrAspTrpThrSerAspAlaValA 907
      |||
4037 ATATCGCCACCCCGGCTTGCTTCAACCGNTACGCGGCGCATTAAG 4086
      ||
907 sn..... 907
      ||
4087 GCAGATTATTCTCAACCGGCGCAACACATNTCCATCAGNCCTTATT 4136
      ::::: |||:::
908 .....LeuGlnPropheal 912
      |||
4137 NAGCGTGTCCATACCGATGCGGCTTCGGGAAAGTCCGAACACGCGTCA 4186
      |||:::|||||
912 aAsnLeuAlaTyrThrHisTyrArgAsnGluGluIleAsnGluGlnGly 929
      |||
4187 ATACCGCGNGTATTG.....GCTCAG 4206
      ::::: |||::: |||:::
929 1yAlaAlaAlaLeuAlaGlyAspLysGlnSerAlaThrAlaSer 945
      |||
4207 GATTTCGGCAAAACCGGATGCGGATGCGGCGTAAACCGCGAAATCAA 4256
      ::::: |||::: |||:::
946 ThrLeuGlyLeuAlaArgAlaAspThrGlnTrpGlnThrAspSerVal..... 960
      |||
4257 AGGTTTACGGGTGTCNTTCACGCTGCGCGCGCAAGNCGCACTGG 4306
      ||| ||| ||| ::::: |||
961 .....AlaIleAlaLeuAlaGly...GluLeuG 969
      |||
4307 AAGCGCAACACAGCGCGGCGC 4326
      |||||::: |||
969 1yTrpGlnHisGlnTyrGly 975
```